

## **2MASS Extended Source Catalog: Overview and Algorithms**

T.H. Jarrett, T. Chester and R. Cutri  
Infrared Processing & Analysis Center, MS 100-22,  
California Institute of Technology,  
Jet Propulsion Laboratory, Pasadena, CA 91125  
jarrett@ipac.caltech.edu  
tchester@ipac.caltech.edu  
roc@ipac.caltech.edu

S. Schneider and M. Skrutskie  
Astronomy & Physics Department,  
University of Massachusetts  
Amherst, MA 01003  
skrutski@phast.umass.edu  
schneider@phast.umass.edu

J.P. Huchra  
Harvard Smithsonian, CfA  
Cambridge, MA 02138  
huchra@cfa.harvard.edu

Received \_\_\_\_\_; Accepted \_\_\_\_\_

for submission to the Astronomical Journal

## ABSTRACT

The 2 Micron All-Sky Survey (2MASS) will observe over one-million galaxies and extended Galactic sources covering the entire sky at wavelengths between 1 and 2  $\mu\text{m}$ . Most of these galaxies, from 70 to 80%, will be newly catalogued objects. The survey catalog will have both high completeness and reliability down to  $J = 15.0$  and  $K_s = 13.5$  mag, or 1.7 mJy and 3.1 mJy, respectively. Galaxies as small as  $10''$  are resolved and as large as  $\sim 2.5'$  are fully imaged. 2MASS will discover galaxies never seen before in the "zone of avoidance" caused by the obscuring effects of Galactic dust and gas, limited only by the extreme number of stars near the Galactic center and at very low Galactic latitudes.

The first public release of 2MASS data occurred in the Spring 1999, containing  $\sim 2500 \text{ deg}^2$  of northern hemisphere sky, including  $> 75,000$  extended sources. For each extended source we report an accurate position, size and shape, flux and surface brightness measures, 2-D image parameters, symmetry metrics (morphology proxies), and include a full resolution "postage stamp" data cube containing the J, H and  $K_s$  imaging data.

This paper describes the basic algorithms used to detect and characterize extended sources in the 2MASS database and catalog. A future paper will provide a full statistical analysis and verification of the completeness, reliability and integrity of the first release catalog, as well as some of the basic scientific results of the catalog.

## 1. Introduction

The Two Micron All-Sky Survey (hereafter, 2MASS) is a ground-based, all-sky survey that utilizes the near-infrared band windows of J (1.13 - 1.37  $\mu\text{m}$ ), H (1.5 - 1.8  $\mu\text{m}$ ) and  $K_s$  (2.0 - 2.3  $\mu\text{m}$ ). Conceived over a decade ago (Kleinmann et al. 1994), the project has evolved from an extensive prototype engineering phase (Beichman et al. 1998) to the current operational phase in which survey data has been acquired and accumulating beginning in the spring of 1997 (Skrutskie et al. 1997). Two dedicated 1.3-m telescopes, one covering northern declinations and one covering southern declinations,

were designed specifically for 2MASS to provide all-sky uniformity. The data acquisition operations are expected to continue up to 2001 when the sky will have been covered >98% with satisfactory photometric precision and uniformity. The first public release of 2MASS data occurred in the late fall of 1998, and the first large incremental release in the spring of 1999.

The point source sensitivity limits ( $10\sigma$ ) are 15.8 (0.8 mJy), 15.1 (1.0 mJy) & 14.3 (1.3 mJy) mag at J, H,  $K_s$ , respectively. The extended source sensitivity ( $10\sigma$ ) is  $\sim 1$  mag brighter than the point source limits, or 14.7 (2.1 mJy), 13.9 (3.0 mJy) & 13.1 (4.0 mJy) mag at J, H,  $K_s$ , respectively. Given the  $\sim 2''$  angular resolution of the image data and the detector sensitivity, 2MASS is well suited for detecting most types of galaxies to  $cz < 10,000$  km/s and high luminosity giant galaxies beyond 30,000 km/s. In addition to galaxies, 2MASS will also identify compact and diffuse Galactic objects. The 2MASS catalogs are expected to detect over 100 million stars and >1 million galaxies (Chester & Jarrett, 1998). This paper will focus upon the detection, identification and characterization of 2MASS extended sources. Future papers will focus more on specific scientific studies with the 2MASS extended source catalog.

The scientific objectives of the extended-source portion of 2MASS include studies of large scale structure, utilization of the infrared Tully-Fisher relation, a complete survey of the local group of galaxies, and an unprecedented census of galaxies located behind the plane of the Milky Way, often referred to as the "zone of avoidance." As such, survey requirements were established in order to satisfactorily achieve these science goals. In addition to the sensitivity limits given above, the extended source Level-1 Specifications include >90% completeness and >98% reliability for most of the sky (free of stellar confusion). There are no set requirements for observations deep in the Galactic plane, but the survey maintains a high level of utility all the way down to  $b = 0^\circ$  (Jarrett et al. 1999).

The level-1 science requirements apply to the galaxy catalog derived from the 2MASS database. The basic 2MASS data and pipeline reduction overview is given in §2, including discussion of the point spread function – a basic component of star-galaxy separation. In §3 we describe some of the key parametric measurements made on extended sources and the crucial operational step of background removal. Sections 3 & 4

describe the algorithms developed to cleanly discriminate between point sources and extended sources. The catalog reliability criterion is in particular a difficult goal to achieve, necessitating implementation of algorithms specifically designed to perform star-galaxy separation with 2MASS imaging data, described in §4. Finally, in §5 we give some examples of the wide array of extended sources that 2MASS is encountering. We will present more detailed scientific results from the 2MASS catalog, including galaxy colors, source counts, completeness and reliability, and clustering in future papers.

## 2. Data and Basic Reductions

### 2.1. Image Data

The 2MASS survey strategy is to map the sky with overlapping strips, or tiles, each of approximately  $6^\circ$  in length and  $8.5'$  in width, using three (one for each band)  $256 \times 256$  NICMOS (HgCdTe) arrays ( $2''$  pixels). The data are efficiently acquired with a freeze-frame scanning technique (detailed in Beichman et al. 1998), such that every piece of sky is observed a total of six times at 1.3 s of integration per sample. With careful sub-pixel dithering between samples, the deleterious effects of under-sampling are minimized. Frames are optimally combined to form “Atlas” images of size  $512 \times 1024$  pixels with resampled  $1''$  pixels. In this paper the Atlas image is also referred to as the “coadd” image. Atlas images have  $\sim 10\%$  overlap along the in-scan (declination) axis to minimize incompleteness of large galaxies. Thus, each  $6^\circ$  scan is comprised of 23 coadd images. The Atlas image is the basic data product from which galaxies and extended sources are detected, characterized and extracted into the 2MASS database. In addition to the full coadd images, small sub-sections of the Atlas images (referred to as “postage stamp” images) are extracted for each extended source.

### 2.2. Pipeline Reductions Overview

High level data reductions include linearity, dark frame subtraction and pixel-to-pixel gain correction (i.e., flat-field correction), which are formed in a non-standard fashion to accommodate the data set unique to the 2MASS survey (see 2MAPPS Functional Design Document 1996; Beichman et al. 1998, Cutri 1988). Further pipeline

reductions include frame-to-frame offset determinations, simple background subtraction, source detection, atmospheric "seeing" and point spread function (PSF) characterization, stellar photometry, band merging, artifact removal, accurate position reconstruction, and photometric calibration. The source detection step described below is vital to both point source processing and extended source processing. The extended source processing occurs at the end of the 2MASS data reduction pipeline. The main objective of the 2MASS extended source processor (referred to as GALWORKS) is to parameterize source detections and determine which sources are "extended" or resolved with respect to the PSF. Consequently, one of the many vital operations for successful star-galaxy discrimination is the accurate measurement of the PSF.

### 2.3. Source Detection

The primary 2MASS source detection procedure is designed to locate both point sources (primarily stars) *and* extended sources (primarily galaxies). The detection thresholds are chosen to assure complete detection of galaxies brighter than the level-1 specification,  $K_s \sim 13.5$ ,  $J \sim 15$ , over a wide range in surface brightness. For fainter low surface brightness galaxies the completeness will steadily fall off with flux, hence a separate detection step is carried out to find these objects (described in §3.3).

The detection algorithm is closely modeled after the DAOPHOT FIND algorithm (Stetson 1990) which was devised to find stars over a wide range of stellar number density. Each Atlas image is convolved with a 4" FWHM Gaussian over a 13 pixel sub-array averaged to zero. The resulting zero-sum filtered image is thresholded at  $\sim 3$  times the estimated noise level for the initial Atlas image, with detections corresponding to each central maximum within a thresholded region. A rough position and flux is estimated from the corrected (convolved image) centroid. The detection list is then fed to a PSF characterization task (see §2.4 below) and finally to a PSF profile-fitting photometry processor, where positions and integrated fluxes are refined. The detection thresholds ( $3\sigma$ ) correspond to  $J \sim 16^{\text{th}}$  mag for point sources; thus typical galaxies thresholds are  $\sim 0.5$  mag brighter, well beyond the extended source requirements. Extended sources are ultimately identified from this inclusive detection source list.

## 2.4. A Generalized Point Spread Function

The first step toward discerning extended sources, including galaxies and Galactic nebulae, from point sources (mostly stars) is to characterize the point spread function (PSF) accurately. The distinctive shape of the 2MASS PSF derives from a combination of factors: the optics, large 2" pixels (frame images), dithering pattern of the six samples that comprise the coadd image, focus, sampling/convolution algorithm to generate the coadds, and atmospheric seeing. As such, the 2MASS PSF corresponding to frame-coadded images is not well fit with a simple gaussian function. Fortunately it is adequately characterized by a generalized exponential function (see below) out to a radius  $\sim 2 \times \text{FWHM}$ , which is all that is required for star-galaxy discrimination.

The 2MASS PSF typically varies on time scales of  $\sim$ minutes due to two effects: atmospheric "seeing" and thermally-driven variable telescope focus. The 2MASS telescopes are designed to be mostly free of afocal PSFs (under most conditions), but 2MASS images can be slightly out of focus during periods of rapid change in the air temperature – conditions that generally only occur during the hottest summer months. Out of focus images have the difficult property of possessing elongated PSFs. Fortunately, under most/typical observing conditions for the survey, the PSFs are symmetric throughout the focal plane. That leaves the atmospheric seeing as the primary dynamic to the radial size of the PSF. Given the long exposure times per sample (1.3 s) and the six-sample coaddition (with optimal dithering to produce round PSFs), seeing changes result in a symmetric "puffing" in and out of the resultant coadd PSF. We can represent the image PSF with the generalized radially symmetric exponential of the form:

$$f(r) = f_0 \exp \left[ - \left( \frac{r}{\alpha} \right)^{1/\beta} \right]$$

where  $f_0$  is the central surface brightness,  $r$  is the radius in arcsec, and  $\alpha$  and  $\beta$  are free parameters. This versatile function not only describes the 2MASS PSF, but it is also used to characterize the radial profiles of galaxies, from disk-dominated spirals ( $\beta$  close to unity) to ellipsoidal galaxies ( $\beta \sim 4$ , de Vaucouleurs law). However, only galaxies brighter than  $K \sim 12$  mag have reliably separate determinations of  $\alpha$  and  $\beta$ . For fainter galaxies, the parameters gradually converge to those appropriate to the PSF. Thus the scale-length,  $\alpha$ , and the modifier,  $\beta$ , are generally quite correlated, so we combine them

to form a “shape parameter,”  $\alpha \times \beta$ . This parameter is a powerful discriminant: galaxies tend to have larger values of both  $\alpha$  and  $\beta$  than stars; hence, the multiplicative join of the exponential fitting parameters amplifies the difference between point sources and extended sources.

Our ability to track the seeing on short time scales depends on the density of stars. The more stars available to measure a statistically meaningful value of the “shape,” the higher the frequency of seeing changes that can be tracked. The stars must be isolated sources free of contamination from other stars and fainter background stars. A reasonable shape value can be derived from a minimum of about 10 stars. Consequently, for low stellar-density regions, like the north Galactic pole ( $\sim 300$  stars per  $\text{deg}^2$  brighter than  $14^{\text{th}}$  mag at K) the seeing is tracked on time scales of about 30 s; for high density regions ( $>10^4$  stars per  $\text{deg}^2$ ) is tracked on time scales of a few seconds of time. Experience has shown that the seeing can indeed significantly change on times scales as fast as a second of time (see below).

The mean “shape” is determined from an ensemble of isolated stars spatially clustered along the 2MASS in-scan direction. The sample population must be free of extended sources (galaxies) and double stars to be a meaningful measure of the PSF. We employ an iterative selection method that is keyed by using an initial boot-strap from the *lower quartile* of the total population histogram. Since isolated stars will have an inherently smaller “shape” value than extended sources (or double stars), the lower quartile (25%) is dominated by isolated stars and conversely, the upper quartile by galaxies. Hence, the distribution's lower quartile serves as a good first guess to the actual mean shape value of isolated stars. Once the lower quartile is identified, we can iteratively search a restricted range in the histogram to arrive at a stable and robust estimation of the true mean shape value for isolated stars. The initial restricted range corresponds  $-3\sigma$  to  $+2\sigma$  of the lower quartile, where  $\sigma$  is the scatter in the “shape” value. In the first iteration we use an a priori determination of  $\sigma$ . For each iteration thereafter, we set hard limits of  $\pm 2\sigma$ . The final “shape” value corresponds to the median (50% central quartile) of the restricted histogram sample, and the  $\sigma$  to the rms scatter or standard deviation of the population.

In this way we build a “stellar ridgeline” of shape values as a function of scan position. Two very different examples are illustrated in Fig. 1 and 2. The plots show the median “shape” values (large filled circles) along the scan. Extracted sources (including stars and galaxies) are denoted with small points. The corresponding FWHM of the PSF (fit with a gaussian) are also shown to give some idea of the scale in arcseconds. In Fig. 1 we show the resultant ridgeline for a scan passing through the Hercules cluster of galaxies. The stellar number density is not large (Galactic latitude of Hercules is about  $30^\circ$ ), but there are still plenty of isolated stars easily separated from extended sources with galaxies located above the mean “shape” ridge. The seeing is fairly stable for each band all throughout the  $6^\circ$  scan spanning  $\sim 6$  minutes of time. The same cannot be said for the second case, Fig. 2, which demonstrates both poor seeing conditions and very rapid changes in the seeing. Fortunately, the stellar density is rather high in this field, 4000 stars per  $\text{deg}^2$ , and the rapid seeing diversions are, for the most part, sufficiently tracked. Scans for which the seeing is poorly tracked or the absolute value of the mean scan seeing is greater than  $1.3''$  (PSF FWHM  $> 4''$ ) are considered low quality data and are in most cases scheduled for re-observation. The stellar ridgelines are used in the extended source processing to separate point sources from "resolved" sources.

## 2.5 Data Anomalies and Artifacts

Data anomalies and image artifacts come in a variety of flavors, including those from the local environment (Observatory and atmosphere), space (meteor streaks, bright stars), from the equipment (array detectors, telescope tracking and focus), and from the software (algorithm and pipeline defects). Extended sources are vulnerable to most of these problems, but in particular those in which the image backgrounds are corrupted. For example, bright stars ( $K_s < 7\text{th mag}$ ) induce several image artifacts, including confusion halos, large-angular extent diffraction spikes, horizontal striping, persistence ghosting, reflection glinting, and large-scale background corruption for the brightest stars ( $K < 3\text{rd mag}$ ). Detection and removal of artifacts is a high priority in the reduction pipeline software and post-processing catalog generation. Still, it is not possible to eliminate all artifacts from the image and source products. Further discussion of data anomalies and artifacts is given in the text to follow.



### 3. Overview of Extended Source Processor

The last major subsystem to run in the 2MASS quasi-linear data reduction pipeline is the extended source processor, referred to as GALWORKS. The primary role of the processor is to characterize each detected source and decide which sources are "extended" or resolved with respect to the point spread function. Sources that are deemed "extended" are measured further and the information is output to a separate table. In addition to tabulated source information, a small "postage stamp" image is extracted for each extended source from the corresponding J, H and K<sub>s</sub> Atlas images. The source lists and image data are stored in the 2MASS extended source database; see flow schematic, Fig. 3.

The 2MASS extended source database contains several classes of "extended objects," including real galaxies, Galactic nebulae and pieces of large angular-size sources, Galactic H II regions, multiple stars (mostly double stars), artifacts (pieces of bright stars, meteor streaks, etc.) and faint (mostly point-like) sources with uncertain classifications. For extended sources, the ultimate goal of the 2MASS project is to produce a reliable catalog of real extended sources, predominantly galaxies. It is therefore necessary for additional "post-processing" steps to eliminate artifacts and confusing objects like double stars. In §4, we discuss in detail how the star-galaxy separation process is performed. For the GALWORKS processor, the emphasis is placed primarily on completeness; that is, we want to comprehensively detect and identify extended sources (especially galaxies) brighter than the level-1 specifications limits of  $K \sim 13.5$ ,  $H \sim 14.3$  and  $J \sim 15.0$ . Later in the post-processing operations phase the galaxy completeness is relaxed (but still within the level-1 specifications) in order to achieve the desired reliability in the galaxy catalog.

2MASS is an all-sky project that will acquire something like ~40 Tb of data over the lifetime of the project. This places severe runtime restrictions on the pipeline reduction software; consequently, one important caveat is that most of the GALWORKS algorithms and flow structures were designed specifically to run and operate as fast and as efficiently as possible, with some functionality omitted toward this end.

By the time GALWORKS is run in the 2MASS pipeline, point sources have been fully measured with refined positions and photometry, band-merged, coordinate positions calibrated, Atlas images constructed, and the time-dependent PSF characterized.

GALWORKS, in addition, generates PSF ridgelines (§2.4) on finer time scales depending on the number of sources available from the scan (i.e., the stellar number density), which are then used to parameterize sources and perform basic star-galaxy discrimination. The high-level steps that encompass GALWORKS include: (1) bright star (and their associated features) removal, (2) large ( $>4'$ ) cataloged-galaxy removal, (3) Atlas image background subtraction, (4) star counting and measurement of the confusion noise, (5) source parameterization and attribute measurements, (6) star-galaxy discrimination (discussed at length in the next section), (7) refined photometric measurements, and finally (8) source and image extraction; see flow schematic, Fig. 4.

The background subtraction operation is a particularly crucial step since both star-galaxy discrimination and photometry rely upon accurate zeroing, smoothing and flattening of the image background. This operation is described in detail below. Steps 4-6 are designed to isolate "normal" galaxies and other relatively high surface brightness extended sources. There are, however, other kinds of extended sources that 2MASS is capable of detecting, including bright Galactic young stellar objects (H II regions, T-Tauri stars, etc.), faint nebulae and low surface brightness. These objects tend to be relatively rare and/or constrained to relatively small angular-sized fields toward the Galactic plane (e.g., molecular clouds) and as such there are no set requirements for their detection completeness or reliability. A separate catalog of bright extended stars and faint LSB galaxies will be released at a later date. A description of the algorithm to detect stars with associated extended emission is described in §3.2 and the algorithm to detect low surface brightness galaxies in §3.3. The remainder of this paper focuses on the detection, characterization and extraction of "normal" galaxies.

### 3.1 Atlas Image Background Removal

In the near-infrared, the background "sky" emission has structure at all size scales, primarily due to upper atmospheric aerosol & hydroxyl emission (the so-called "airglow" emission; Ramsey et al 1992). The OH emission is the dominant component to

the J (1.3  $\mu\text{m}$ ) and H-band (1.7  $\mu\text{m}$ ) backgrounds, while thermal continuum emission comprises the bulk of the K (2.2  $\mu\text{m}$ ) background. The J and H images tend to have more background “structure,” and at times of severe airglow the background can have high frequency features on scales of tens of arcseconds that can trigger false extended source detections. For extended sources, the primary objective of the 2MASS project is to find and characterize galaxies (and other extended objects) smaller than  $\sim 3'$  in diameter. We therefore attempt to remove airglow features slightly larger than this limiting size scale to minimize random and systematic photometric error from non-zero background structure. For the most part, the background variation in a given image (size  $8.5 \times 17'$ ) is smooth and can be modeled with a polynomial. A third order polynomial turns out to be a good compromise between a simple planar fit and a series of spline waves.

For the case in which the airglow frequency of variation is higher than we can adequately remove, the resultant photometry (particularly at H band) is severely compromised. These data are given a lower quality score and are scheduled for re-observation if time permits.

Each Atlas image is separated into three  $512 \times 512$  pixel blocks (see Fig. 5) for the purpose of background fitting. Using a least-squares technique, a cubic polynomial is iteratively fit with  $3\sigma$  rejection to each line of the after  $8 \times 8$  pixel median filtering. The line solutions are used for input to the next step, where we fit a cubic polynomial to each column, thereby coupling the line and column background solutions. The three block solution images are averaged with a  $(1/\Delta r)$  taper. Here  $\Delta r$  refers to the relative radial (“in-scan”) difference between any two given block solutions from the middle of each block at  $Y = 256, 512$  and  $768$ . So for example, combining the lower and central blocks at some point,  $Y'$ , gives the respective weights  $[1 / (256 - Y')]$  and  $[1 / (512 - Y')]$ . With this technique we are able to smoothly combine the three independent solutions per coadd image. Note however, the boundary solutions for the upper and lower blocks are better constrained near the center of the image due to the weighted addition of the central block solution image. Conversely, the background solutions are not as well determined at the upper,  $>900$ , and lower,  $<124$ , “in-scan” image extremes. The fitting schematic is illustrated in Fig. 5. The  $512 \times 1024$  coadd is represented by a thick-lined rectangle. As explained above, cubic fits are applied to the lower half,  $512 \times [1:512]$  pixels, the upper

half,  $512 \times [513:1024]$  pixels, and the central half,  $512 \times [257:768]$  pixels, where we have first resampled the data with an  $8 \times 8$  median filter.

The background removal process is applied separately to the J, H, &  $K_s$  coadd images ( $512 \times 1024$  pixels each). Given the “cross-scan” size of one coadd image, a cubic polynomial,  $ax^3 + bx^2 + cx + d$ , provides an effective model for smooth background variations larger than  $\sim 3'$ . In the “inscan” direction the larger size ( $1024''$ ) allows cubic fit to each half of the coadd (lower  $512''$ , upper  $512''$ ), and we also apply a fit to the “central”  $512 \times 512$  pixels in order to smoothly “join” the boundaries of the two background solution fits. The final  $512 \times 1024$ -solution fit is generated from a weighted average of each  $512 \times 512$ -block solution. The fitting procedure is first preceded by an image “clean” operation. Stars and catalogued galaxies are masked from the image. Very bright stars ( $K < 6$ th mag) require more complicated masking, including removal of their bright internal reflection halo, diffraction spikes, horizontal streaks, filter glints and persistence ghosts. Finally, in order to minimize contamination from faint stars and objects that escaped the masking procedure, we median clean the Atlas image with an  $8 \times 8$  pixel filter (thus, degrading the resolution of each pixel to  $8''$  chunks).

Representative performance of the background removal operation is shown in Fig. 6. The image data comes from a fairly typical “photometric” Northern Hemisphere night, although the “airglow” emission is fairly severe during the period that this data was acquired (see H-band, middle panels). The figures show the raw image coadd, resultant background solution and residual (background subtracted) image. The gray-scale stretch ranges from  $-2\sigma$  to  $5\sigma$  of the mean background level. The J, H & K raw images reveal fairly low level (smooth, but non-linear) background variations, while the corresponding residual images show very little (if any) background structure. However, airglow emission is much more prevalent in the H-band, with size scales smaller than  $\sim 1-2'$ , as evident in the residual image. It is this residual structure in the background (with amplitude  $> 10\%$  of the mean background noise) that can induce systematics in the photometry, parameterization (e.g., azimuthal ellipse fitting), and reliability. In the future we will attempt to apply higher order fits (e.g., using splines) in an attempt to fit out these airglow features. The projected success of this procedure is unknown at this time.

### 3.2 Bright Extended (Fuzzy) Stars

Bright fuzzy stars are identified using a separate algorithm within the GALWORKS pipeline (Fig 4). The basic method is to look for emission in and around the source at levels elevated above that expected for a bright star characterized by the PSF. After nearby stars have been masked and the source itself has been subtracted based on the shape of the PSF, the remaining or residual emission is measured with respect to the mean background level of the coadd. A more effective indicator of enhanced emission is to calculate the root mean square of the residual emission versus the mean background AND versus a zero background (i.e., assume the true background level is zero). The rms values are then normalized by the measured noise for the coadd (Atlas image) as a whole. Stars with associated emission, like reflection nebulae, clearly stand out. This operation is referred to as the “bright extended source” processor. Extracted sources are stored in the 2MASS database and a special catalog to be released at some date in the future. There are no set requirements for these kinds of objects and the completeness and reliability of this supplemental catalog are unknown at this time. Examples of sources found with this technique are shown in Fig. 7, from scans crossing the Orion trapezium and the Large Magellanic Clouds. The top row shows J-band “postage stamp” images, middle row the H-band and bottom row the  $K_s$ -band images. Each image is  $50''$  in width. The integrated flux for the example sources range from  $5^{\text{th}}$  to  $7^{\text{th}}$  mag.

### 3.3 Low Central Surface Brightness Galaxies

Low surface brightness galaxies (e.g., dwarf ellipticals among other types) present a different challenge to GALWORKS than the typical “normal” galaxy 2MASS encounters. They are generally very faint (as measured in a standard aperture for “normal” galaxies) and they do not have well defined cores; see Fig. 8 for examples of typical low central surface brightness galaxies found within 2MASS. Each image is  $25''$  in width. The integrated flux of the example sources range from  $J=15$  to  $15.6$  and  $K_s=13.8$  to  $15.1$ . The LSB galaxy nature of many of these sources was confirmed with

deep optical images. There are some examples of galaxies observed to be low surface brightness in the near-infrared but normal in the optical—typically blue spiral galaxies.

The galaxy core is an important component for star-galaxy separation since many of the parametric measurements for star-galaxy separation are anchored to the core of the galaxy. The low central surface brightness detector (referred to as the LCSB processor) of GALWORKS is executed last in the chain of operations that comprise GALWORKS (see flowchart). The input to the LCSB processor is a fully cleaned coadd image in each band, where stars (brighter than some limit, typically  $K_s = 14.5$ ) and galaxies have been entirely masked. The image is then blocked up (using three independent kernel sizes:  $2 \times 2$ ,  $4 \times 4$  and  $8 \times 8$ ) and "boxcar" smoothed to increase the signal to noise ratio for large (but faint) objects normally hidden in the  $1''$  pixel noise. A block average is not the optimum method (as compared to a gaussian convolution, for example) but with pipeline runtime constraints it is the best option.

The detection step consists of  $3\text{-}\sigma$  threshold isolation of local peaks in the blocked-up cleaned images. Source detections are then parameterized (using the blocked and smoothed image) with the primary measurements being: signal to noise ratio of the peak pixel, radial extent (SH score), integrated signal to noise, surface brightness, integrated flux, and SNR measurements using a  $J+H+K_s$  combined "super" image. The "super" image, in principle, provides the best median from which to find faint LSB galaxies given the effective increase in the signal to noise ratio. In practice, the "super" image only increases the SNR by approximately 30% due to the significant background levels at H and K and assuming normal (i.e.,  $J-K \sim 1$ ) galaxy colors. Faint stars remaining in the cleaned image will have a relatively low SNR since most of their light is confined to a few pixels that are averaged with blank sky in the blocking and boxcar-smoothing step. Galaxies, on the other hand, will add up since their light is distributed over a large area.

The preliminary results for the LCSB processor reveal a reliability rate of about ~70 to 80% using a threshold on the "maximum" SNR (between  $2 \times 2$ ,  $4 \times 4$  and  $8 \times 8$  blockings) of the "super" coadd. The major contaminants are faint stars and diffuse emission associated with bright stars. However, if a meteor streak (or other transient phenomenon) occurs in the coadd image, then numerous false sources are picked up as

LSB galaxies. It remains to learn how to improve the reliability of sources coming from the LCSB detector. It is important to note that these sources are nearly always fainter than the level-1 specifications ( $K_s > 13.5$ ,  $J > 15$ ) which means that there are currently no requirements on the incompleteness and reliability. We do not anticipate significant completeness failure for LSB galaxies brighter than  $K_s \sim 13.5$ . The fainter LSBs, however, will have to be detected and processed with the LCSB processor described here and released in a future special catalog. Further information and some early science results with 2MASS LSB galaxies can be found in Jarrett (1998) and Schneider et al. (1998).

### 3.4 Initial Source Parameterization

Preliminary flux estimates come from the point source processor, which uses a characteristic PSF to derive total fluxes (assuming a point-like flux distribution). These measures systematically underestimate the flux of extended sources. Hence, one of the first tasks for GALWORKS is to deduce the nature of a source using some simple radial profile attributes. The median radial shape score, or MSH, is both quick to compute and a robust discriminator between stars/double stars and galaxies (see §4.1 for a complete definition of a star-galaxy "score"). Applying an extremely conservative threshold to the MSH measure for each source in each band separately eliminates a large fraction of the total number of sources that require more exhaustive testing for star-galaxy separation. If the source is very likely to be extended (large MSH score), then its flux is re-estimated using a fixed  $R=10''$  circular aperture.

Before the more time-consuming image attribute measurements are performed on each source (e.g., elliptical shape fitting; adaptive aperture photometry), it is necessary to perform additional star-galaxy separation tests, particularly when the stellar number density is very high, as at  $l \leq 10^\circ$ . Thresholds on the SH, WSH, R1, and R23 radial shape attributes (see §4.1) are carried out to eliminate additional non-extended sources (namely stars and double stars) from the source list. For high latitude fields, the remaining sources (in a typical scan) are mostly real galaxies intermixed with a few double stars, one or two isolated stars and low S/N objects of uncertain nature. The

reliability is from 50 to 80% at this juncture, and thus the star-galaxy separation process has reduced the fraction of stars to galaxies from 10:1 to approximately 1:1.

### 3.5 Ellipse Fitting

The ellipse-fitting method was designed to run fast and to minimize confusion from nearby sources (i.e., stars) and correlated noise features that form "extended" limbs and other disconnected extended features. Three elliptical parameters are derived from the image isophote: axis ratio ( $b/a$ ), position angle,  $\phi$  (standard reference frame, east of north), and a goodness of fit metric. The goodness of fit is defined as follows:

$$\min \left( \frac{\Delta r_{\text{semi}}}{\bar{r}_{\text{semi}}} \right)$$

where  $r_{\text{semi}}$  is the semi-major axis corresponding to some point along the isophote and a given (axis ratio,  $\phi$ ) solution,  $\Delta r_{\text{semi}}$  is the population standard deviation and  $\bar{r}_{\text{semi}}$  is the population mean. That is to say, for a given solution ellipse (described by the axis ratio and position angle) the resultant shape of the distribution of  $r_{\text{semi}}$  along the isophote tells us the goodness of the fit. If the ellipse ( $b/a$ ,  $\phi$ ) is perfectly matched to the isophote, the mean variance in  $r_{\text{semi}}$  is identically zero. If the match is poor, then the variance is large while the population mean can be large or small. Therefore, by minimizing the ratio of the standard deviation to the mean radius in the distribution, we arrive at the best-fit ellipse solution. In this fashion, the elliptical parameters are derived for each band.

An additional fit is performed on the combined (J+H+K<sub>s</sub>) "super" coadd image. The "super" coadd has a higher signal-to-noise ratio than the individual fits. Accordingly, the derived "super" coadd ellipse serves as the "default" shape for cases in which the individual band flux is fainter than:  $\sim 14.4$  at J,  $\sim 13.9$  at H and  $\sim 13.5$  at K<sub>s</sub>, or the SNR of the galaxy is less than 5, based on the  $R=10''$  fixed circular aperture photometry. For the case in which the derived semi-major axis is less than  $5''$  or greater than  $70''$ , the source is assumed to be round and the parameters are set accordingly. For the case in which the derived axial ratio is less than 0.10, the ellipse fit parameters are set to the corresponding fit from the "super" coadd. Finally, the "super" coadd values are also used when the individual band fit for one reason or another is not possible (e.g., when masked pixels are present within  $1''$  of the peak pixel).



The orientation of disk spiral and spheroid elliptical galaxies is estimated using a 2-D ellipse fit to a single isophote surface, which is used to compute various forms of aperture photometry (e.g., Kron, isophotal, etc.) and symmetry parameters used for star-galaxy separation. Although galaxies can change orientation (e.g., ellipticity and position angle) with radius, the 2MASS undersampling and runtime constraints limit fitting an ellipse to a single isophote in each band, which is scaled to estimate elliptical isophotal diameters at the isophotes. Moreover, most 2MASS galaxies are small in size ( $<15''$ ), so for our  $\sim 2''$  angular resolution multiple fits are not especially useful. It is to our advantage that in the near-infrared most galaxies appear to have somewhat more consistent orientations and axis ratios at different radii. To minimize the effect of PSF elongation, the fiducial isophote is fit at roughly a  $3\sigma$  level, typically corresponding to surface brightness  $\sim 20.09$  mag/arcsec<sup>2</sup> at J,  $\sim 19.34$  mag/arcsec<sup>2</sup> at H and  $\sim 18.55$  mag/arcsec<sup>2</sup> at  $K_s$ .

A final note regarding the ellipse fitting operation relates to stellar masking. Bright galaxies ( $K_s < 12.5$ ) in which the inclination is large ( $>40^\circ$ ), are apt to be interpreted as multiple point sources by the initial source processor. Consequently, we do not perform any stellar masking or subtraction specific to the ellipse fitting step, *except* when the stellar number density is high ( $>2000$  stars deg<sup>-2</sup> for  $K_s < 14$ ) in which case it is more favorable to mask out nearby stars given the high probability of contamination. This ellipse-fitting detail should not be confused with the general GALWORKS procedure of near-neighbor masking prior to photometry or symmetry measurements.

Once the general orientation of the galaxy is derived, various "symmetry" measures are evaluated. The radial/azimuthal symmetry of an object is a good indicator of its true nature. Double stars appear asymmetric across the minor axis—since the ellipse is centered on the primary component of the double star. This is also generally the case for triple stars, although there are configurations of  $\geq 3$  stars in which the alignment is symmetric across both the minor and major axes.

One way to measure the "symmetry" of an object is to perform a bi-symmetric spatial autocorrelation. Divide the object across the minor axis. The integrated flux in each half gives the gross bi-symmetric flux ratio. Rotate one side 180 degrees with respect to the other and multiply the resultant pieces. The autocorrelation is then

normalized by the original galaxy (squared). To minimize the effects of noise and the shape of the PSF, very low SNR points ( $< 1.5$ ) and the inner  $3''$  core are avoided in this procedure. In addition to the autocorrelation, we also compute bi-symmetric cross-correlation reduced chi-square,

$$(1/N) \sum \left( \frac{(p - p^*)^2}{2 \times \sigma^2} \right)$$

where  $p$  and  $p^*$  are the points  $180^\circ$  apart that are being compared,  $N$  is the number of points being compared, and  $\sigma$  is the pixel noise. This  $\chi$  measure has the advantages that it has a distribution that is well understood statistically with tabulated confidence ranges, there are no asymmetries in the distribution like those introduced in a ratio comparison, and it is insensitive to low SNR or data points near zero.

Finally, we perform an ellipse fit to the  $5\sigma$  isophote (per band and “super” coadd). Comparison between the default  $3\sigma$  and  $5\sigma$  fit parameters may indicate either a real asymmetry due to stellar contamination or orientation changes as a function of radius. Likewise, the goodness of fit metric can indicate both problems with the fit (due to stellar contamination or noise in the case of faint sources) or real asymmetry in the object.

### 3.6 Photometry

Given the assorted shape, size and surface brightness that galaxies exhibit in the near-infrared, a corresponding diverse array of apertures is used to compute the integrated fluxes. Contamination from stars within or near the aperture boundary is minimized with pixel masking—but still remains significant when the confusion noise is high. Flux from masked pixels is “recovered” with isophotal substitution, where the mean value of the elliptical isophote (based on the elliptical shape parameters,  $b/a$  and  $\phi$ ) replaces the given masked pixel that the isophote passes through. More detailed discussion of stellar contamination and rectification thereof in 2MASS galaxy photometry can be found in Jarrett et al. (1996).

The simplest, and therefore most robust, measures come from fixed circular apertures. Fluxes are reported for a set of fixed circular apertures at the following radii: 5, 7, 10, 15, 20, 25, 30, 40, 50, 60, &  $70''$ , centered on the peak pixel based on the J-band image. We report both the integrated flux within the aperture (with fractional pixel

boundaries) and the estimated uncertainty in the integrated flux. The magnitude uncertainty is based solely upon the measured noise in the Atlas image, which includes both the read-noise component and background Poisson component, as well as the confusion noise component which becomes significant when the stellar source density is high. The uncertainty does not incorporate real errors due to source contamination, background gradients (e.g., airglow ridges with a higher spatial frequency than the background removal process can handle; see §3.3), zero-point calibration error, and uncertainties in the adaptive apertures (e.g., isophotal photometry, see below). A more detailed discussion of the 2MASS galaxy photometry error tree can be found in Appendix A. Contamination or confusion flags are also attached to each flux.

For the great majority of faint galaxies in the 2MASS catalog, small fixed circular apertures give the best compromise between increasing noise due to confusion and missing flux in the faint outer parts of galaxies. In particular, the  $R=7''$  aperture appears to have the optimum match with the coupling between the 2MASS undersampling and PSF elongation, with the H and K background noise, and with the size of galaxies fainter than  $K_s \sim 13$  mag.

Adaptive aperture photometry includes isophotal and Kron metrics. The isophotal measurements are set at the 20 mag per arcsec<sup>2</sup> isophote at  $K_s$  and the 21 mag per arcsec<sup>2</sup> at J, using both circular and elliptically shape-fit apertures. Kron aperture photometry (Kron 1980) employs a method in which the aperture is controlled/adapted to the first image moment radius. The Kron radius, which is frequently used in galaxy photometry as a “total” measure of the integrated flux (Koo 1986; Bertin & Arnouts 1996), turns out to roughly correspond to the 20 mag per arcsec<sup>2</sup> isophotal radius under typical observing conditions. The minimum radius is set at  $R=7''$  due to the rapidly increasing (PSF shape and background noise) uncertainty in the isophotal or Kron radial measurement for radii smaller than this limit.

For purposes of computing colors, two classes of adaptive photometry are carried out: individual and fiducial. “Individual” photometry refers to the use of a different adapted aperture per band, which is useful for single-band limited studies. However, the real power of 2MASS data is having simultaneous J- $K_s$ , J-H and H- $K_s$  colors. Colors require a consistent aperture size and shape for all three bands, based on either the J or  $K_s$

isophotes, respectively referred to as the "J fiducial" and "K fiducial" photometry. For the brighter galaxies in the catalog,  $K_s < 13$  mag, the "K" fiducial isophotal elliptical aperture photometry appears to give the most precise measurement (based on repeatability tests), but errors in the ellipse fit to the isophote (see §3.2) result in an uncertainty that is difficult to evaluate (see Appendix A). The adaptive circular apertures reduce some of that uncertainty, but do increase the overall noise due to additional sky noise within the non-optimized aperture—resulting in a less precise, but more robust measurement.

Additional flux measures include the central surface brightness (peak pixel flux) and the “core” surface brightness (average flux over a 5" radius). Finally, a “system” measurement is carried out in which no stellar masking is performed, nor any masking of flux from neighboring galaxies. The “system” flux indicates the total flux in and about a galaxy, so it will include the total light in closely interacting systems. A set of contamination flags supplement the system measurements: one indicating stellar contamination and the other neighboring galaxy "contamination."

A cautionary note: like all isophotes used in 2MASS pipeline processing, the magnitudes are uncalibrated and may be adjusted by  $\sim 0.1$  to  $0.2$  mag in the later calibration processing step. Consequently, the isophote at which the 2-D elliptical parameters are derived can vary from (in background noise units)  $\sim 2.6\sigma - 3.7\sigma$ , depending on the calibration correction.

### 3.7 Source Positions

In addition to the coordinate position based on the PSF-fit operation, two additional “extended source” positions are computed. The first is based upon the peak pixel from the J-band image. The second is based upon the intensity-weighted centroid of the J+H+K<sub>s</sub> “super” coadd image. The precision of the peak-pixel coordinate is limited by the 2" resolution and convolution method used to construct/resample Atlas images from raw frames. Based on repeatability tests these positions possess an rms uncertainty of  $\sim 0.5''$ . The “super” centroid coordinate position is more precise since it applies a 2-D centroid to higher SNR data, but it can be more influenced by unusual

morphologies. The estimated uncertainty of the “super” centroid position is  $\sim 0.3$  to  $0.5''$  for normal surface brightness galaxies.

### 3.8 Source Extraction

Sources that pass the star-galaxy discrimination tests and have an integrated flux brighter than the mag limits:  $J = 15.5$ ,  $H = 14.8$ ,  $K_s = 14.3$  MINUS the confusion noise (which is 0 mag for low stellar number density fields), are extracted to the 2MASS extended source database. In addition to the parameters described in previous sections, the source information includes various flags indicating stellar contamination, cross-identification (with previously catalogued *large* galaxies derived from the NASA Extragalactic Database) and processing status. A list of the “standard” extended source parameters can be found in 2MASS Extended Source Explanatory Supplement.

For each extended source, a small “postage stamp” image is clipped from the larger background-subtracted Atlas image. The stamp images are stored in J, H and  $K_s$  fits-format data cube files. The image size is constrained by the final Kron or isophotal radius, with a minimum diameter of  $21''$  and a maximum diameter of  $101''$ . The dynamic image size reflect the limitations of the finite storage capability of the 2MASS database. The stamp image headers provide all of the information needed to extract photometry, positions, etc., except the larger-area environment that was used to remove a local background (§3.1) and evaluate contamination.

## 4. Star – Galaxy Discrimination

The ability to separate real extended sources (e.g., galaxies, nebulae, H II regions, etc.) from the vastly more numerous stars detected by 2MASS is what fundamentally limits the reliability of any extended source catalog. Single isolated point sources represent the purest and easiest construct from which extended sources must be distinguished. More complicated constructs include “double” stars and “triple+” stars, these are generic labels that include both physically-associated multiple systems and chance superposition of stars on the sky. The permutations and combinations of multiple-star characteristics (radial separation, flux difference, color difference, etc.) make them a challenge separate from real galaxies. The surface density of stars and

galaxies is illustrated in Fig. 9. Sky-projected doubles are less than  $\sim 2\%$  of the total stellar counts at high Galactic latitudes, but begin to dominate the total numbers for  $|b| < 5^\circ$ . Even at moderate stellar number density, double stars are comparable in number to galaxies for typical 2MASS flux levels.

There are many competing methods for separating stars from galaxies, from the simplest “CART” methods (i.e., linearly measuring one attribute versus another), to the more sophisticated Bayesian-based methods (e.g., FOCAS; see Valdes 1982), decision trees (Weir, Fayyad & Djorgovski, 1995) and neural networks (Odewahn et al. 1992). Each method was designed in response to increasingly more complicated data sets. For 2MASS, we were faced with the rather unique combination of near-infrared imaging and under-sampled data and variable PSF shape that called for a special adaptation of these procedures.

Early experimentation with existing algorithms (e.g., FOCAS) were unsatisfactory due primarily to the severely undersampled 2MASS PSF, which changes over time scales of minutes. A critical issue for GALWORKS is to accurately measure and track the time-varying PSF while applying some simple CART-like rules to cull out most of the multiple stars and artifacts that mimic real extended sources. The resultant extended source database is approximately 80% reliable for most of the sky. In a post-processing phase, further refinements, including more complicated attribute combinations and decisions trees, are used to produce the extended source catalog at a reliability of greater than 98% for  $K < 13.5$ . Below we describe and discuss some of the more critical parametric measurements and decision tree operations utilized to that end.

#### 4.1 Stellar Ridgelines and Basic Object Characteristics

Resolved sources are initially identified by comparing their radial profiles with that of the nominal point spread function. As is the case for all ground based observations, the PSF changes with time due to the changing thermal environment and dynamic atmospheric “seeing” (see §2.4); additionally, the PSF has an intrinsic spread caused by the pixel undersampling and dither pattern. Both effects are measured and tracked using our generalized exponential function (Eq. 1) and stellar ridge profiles (Fig.

2). The radial “shape” ( $\alpha \times \beta$ ), or simply SH, of a source is compared to the stellar ridge value,  $sh_0$ , and a N-sigma “score” is computed as:

$$\frac{sh(t) - sh_0(t')}{\Delta sh_0(t')}$$

where  $sh_0(t')$  and  $\Delta sh_0(t')$  denote the time variable ridgeline value and its associated uncertainty and  $sh(t)$  the source value, with time  $t'$  as close to real  $t$  as possible. The PSF ridgeline value is stable over all flux levels, so only one value is needed per time interval.

The SH uncertainty includes both measurement error and the intrinsic PSF spread. However, since  $SNR > 10$  stars are plentiful in most areas, the measurement error is minimal compared to the real spread in the PSF. The uncertainty represents the RMS in the SH distribution, but the distribution has triangular-shaped wings (i.e., the scatter in SH falls off linearly) due to the undersampling and sub-pixel dither. Consequently, stars will not have SH values above a threshold of  $\sim 2 * \Delta sh_0$ , but galaxies and other relatively “extended” objects (e.g., double stars) will have scores  $> 2$ . In Fig. 10 we display the J-band “shape” scores of three kinds of objects that 2MASS commonly encounters: stars, multiple stars (double stars and triple+ stars), and galaxies. Stars occupy a locus about zero SH score (essentially defining the ridgeline), while multiple stars lie well above the ridgeline along with galaxies and other “fuzzy” sources. Note that the number of stars displayed has been reduced by a factor of 10 relative to the other plots. The SH score is very effective at separating isolated stars from galaxies at flux levels as faint as  $\sim 15.4$  in J band.

Other GALWORKS-derived image parameters that are effective at separating isolated stars from galaxies include the 1<sup>st</sup> and 2<sup>nd</sup> intensity-weighted moments, ratio of the central surface brightness to the integrated brightness, and areal measures (e.g., isophotal area). Unfortunately, like the radial SH parameter, none of these diagnostics can discriminate galaxies from sky-projected star clusters to the degree necessary to meet the level-1 specifications. Double stars are particularly vexing due to their sheer numbers at  $l|b| < 20^\circ$  (Fig. 9). Double stars (and triple stars near the Galactic plane) are clearly the primary contaminant of the galaxy database. More intricate attributes are needed to exploit the differences between groupings of point sources and genuinely extended sources.

## 4.2 Multiple Star – Galaxy Separation using Symmetry Metrics

In the near-infrared, galaxy morphology usually has smooth radial and azimuthal profiles. Spiral galaxies have much more even light distributions in the near-infrared than optical because the absorption is greatly reduced and the emission is dominated by older stellar populations, including low mass dwarfs and red giants, which are less concentrated in spiral arms. Features commonly seen at the radio and optical wavelengths, including H II regions, supernova remnants and dust lanes, are generally difficult to detect in the near-infrared except in the nearest galaxies; Fig. 11 shows a few large angular scale galaxies located in the Virgo cluster. Only the relatively rare cases of galaxies subject to strong tidal or hydrodynamical interactions exhibit significant asymmetry in the near-infrared bands. In contrast, multiple stars, and in particular double stars, are not symmetric about their "primary" center. Here the center of a multiple star corresponds to the brightest member in the group, or more specifically, the peak pixel associated with the brightest star. The near-infrared symmetry of galaxies can be exploited to differentiate between multiple stars that otherwise mimic extended sources.

Fig. 12 illustrates a variety of double stars seen in 2MASS images. For comparison, a set of galaxies of approximately the same integrated brightness as that of the double stars is also shown in the lower panels. For double stars, the fainter star ("secondary" component) breaks the symmetry about the primary. Hence, the signature of a double star is an asymmetric azimuthal profile, which may be detected in the symmetry measurements discussed in the last section (e.g., flux ratio across the major axis).

A different tactic is to "remove" the secondary and measure the resultant SH of the primary. We are of course faced with the problem that the emission from both sources are entangled and the primary itself has changed both its radial (SH) width and its azimuthal (symmetry) shape. If the PSFs were exceptionally stable and well characterized as such, then in principle it would be possible to satisfactorily de-blend the multiple sources into their constituent parts. Since this condition is not always realized, and moreover the runtime for this kind of multiple PSF  $\chi^2$  fitting is prohibitively long, we are left with less ideal methods. The simplest approach is to remove the secondary using



a median filter in annular shells about the primary: GALWORKS refers to the resultant measure as the “median shape” or MSH. Another more complicated approach is to mask the secondary and measure the residual emission from the primary, using a  $45^\circ$  wedge or pie-shaped mask that is rotated about the vertex anchored to the primary. The optimum configuration in which the secondary is effectively masked is found by rotating the wedge mask through all angles (Fig. 13). The SH score is then computed for the remaining ( $360^\circ - 45^\circ$ ) pixels. If the secondary star is masked, then the resultant SH score will be minimized, ideally with a value corresponding to an isolated star. In practice the secondary can never be fully masked, and the peak pixel does not represent the true center of the primary since it is slightly shifted toward the secondary—thus resulting in a slightly inflated SH score relative to that of an isolated star. Nevertheless, the “wedge” shape score, or WSH, is an effective discriminant. This is as demonstrated in Fig. 14, which is analogous to Fig. 10; here we show the distribution of multiple stars and galaxies as measured in the WSH versus magnitude plane.

The wedge shape score for double stars is considerably smaller than the corresponding SH score, having values typically less than 5 for  $J < 15$ , while galaxies remain “extended” in this measure with scores  $>5$  for  $J < 15$ . Note however, triples+ stars are only occasionally identified as such by the WSH score since the additional two secondary components usually defeats the single rotating mask method. For triple stars, yet more severe “symmetry” constraints are required.

Triple stars are geometrically more difficult to characterize because of the number of possible combinations of integrated flux and primary-secondary separations. For most triple stars there is minimal contamination from the two secondary components along some direction from the primary. If we measure the radial SH of this vector and compare it to the corresponding ridgeline value, the resultant “score” should be close to that of an isolated star. Thus the basic method is to measure the SH along an azimuthally distributed set of vectors at angular separations of  $5^\circ$ . The vector corresponding to the “minimum” shape score (referred to as the R1 score) is susceptible to background noise fluctuations since we are restricting the  $(\alpha, \beta)$  fitting operation to less than a dozen pixels. For galaxies, the R1 score tends to select against galaxies that are edge-on and thus have

minimal (but still measurable) extended emission along the minor axis (i.e., the vector corresponding to the minimum radial SH score).

A more robust parameter (but slightly less effective at removing the influence of the secondary components) is to average the 2<sup>nd</sup> and 3<sup>rd</sup> lowest SH value vectors (that is, avoid the R1 vector). This score is referred to as the R23 shape score. Here we are relying upon the fact that most triple star configurations (but not all by any means) will have more than one vector that is minimally affected by the secondary components. Galaxies, meanwhile, are generally extended in all directions and so the R23 score is not much different from the SH score except for the faintest galaxies ( $J > 15$ ,  $K_s > 13.75$ ) which are at the mercy of noise fluctuations.

The effectiveness of the R23 score is demonstrated in Fig. 15. Here we plot the R23 versus magnitude phase space. It can be seen that the triple stars are now well under control with minimal loss to the galaxies at  $J < 14$ , while for the faint magnitude bins,  $J > 14$ , galaxies are not well separated from triple stars. But, as it turns out, triple stars are only abundant when the stellar number density is very high (i.e., the Galactic plane; see Fig. 9), which means that the "confusion" noise is also high (that is, the random fluctuations in the background due to faint stars; see Appendix C), rendering the sensitivity limits for galaxy detection itself from 0.5 to nearly 2 mags brighter than the high-latitude 2MASS limits. Thus, just as the problem with triple stars becomes significant, the practical detection thresholds are correspondingly decreased, thereby leaving the R23 score as an effective star-galaxy discriminator for flux levels up to the detection limits. For the most extreme stellar number density cases (e.g., regions of Baade's windows),  $>10^5$  stars per  $\text{deg}^2$  brighter than 14<sup>th</sup> at  $K_s$ , quadruple ++ stars become significant, at which point there is little that can be done to separate galaxies from clusters of stars.

We have developed additional parameters designed to discriminate triple stars from extended objects, including measuring the linear flux gradient along radial vectors and the integrated flux gradient along radial "column" vectors (referred to as the VINT score). Similar to the R1 and R23 scores, these methods rely upon the "minimum" column integrated flux or gradient in the column flux to be similar to that of isolated stars. They are not quite as effective as the SH vector scores, but since they are only

slightly correlated, they can be used in combination with the other attributes when using a decision tree classifier.

#### 4.3 The Color Attribute

Two effects conspire to make galaxies appear “red” in the 1-2  $\mu\text{m}$  window: their light is dominated by older and redder stellar populations (e.g., K and M giants), and their redshift tends to transfer additional stellar light into the 2 $\mu\text{m}$  window (for  $z < 0.5$ ), boosting the K-band flux relative to the J-band flux. The latter phenomenon is often called the “K correction,” although the “K” here is unrelated to the infrared band. Because of this, the J-K color attribute can be used—in conjunction with color-independent discriminants, like the WSH score—to cleanly separate extragalactic objects from stars. As a bonus, the color separation is enhanced in the Galactic plane where double and triple star contamination is severe. This is because galaxies are subject to a larger column dust compared to random field stars along the same line of sight. In Fig. 16 we demonstrate the effectiveness of the J-K color to separate stars from resolved galaxies in a diverse set of fields, including areas well above the Galactic plane, referred to as low stellar density fields ( $< 10^{3.1}$  stars per  $\text{deg}^2$  brighter than  $14^{\text{th}}$  at K), and areas closer to the plane ( $|b| > 5$  degrees), referred to as moderate density fields ( $< 10^{3.6}$  stars per  $\text{deg}^2$ ), and finally areas in the Galactic plane in which the stellar number density is very high ( $> 10^{3.6}$  stars per  $\text{deg}^2$  brighter than  $14^{\text{th}}$  at K). For the latter case, the confusion noise is typically very high ( $> 1$  mag) so the sensitivity limits have been decreased accordingly.

A J- $K_s$  color of  $\sim 1.0$  appears to be a reasonable compromise for separating stars from galaxies. For flux levels relevant to the 2MASS level-specifications,  $K_s < 13.5$ , a J- $K_s$  color limit of 1.0 eliminates nearly all ( $> 95\%$ ) double stars that mimic galaxies, while more than 90% of the total galaxy distribution has a color greater than this limit.

Another way to view the color separation between stars and galaxies is within the J-H vs. H- $K_s$  color plane, Fig. 17. Here we include the stellar main sequence track, showing the divergence of giants from dwarfs at  $H-K_s > 0.3$ . In addition, we show the K-correction track for spiral galaxies derived from the models of Bruzual & Charlot (1993). When the surface density of stars is high the extinction is also on the rise, clearly seen in the right panel of Fig. 17.

At fainter flux levels,  $K_s > 13.5$ , the scatter in the integrated flux (and thus colors) is large enough that false galaxies (i.e., double and triple stars) can scatter above the J- $K_s$  color limit and galaxies can have colors that scatter below the limit to a degree that contamination and completeness is significantly compromised if the J- $K_s$  attribute were used as the lone discriminant. Moreover, for all flux levels, a J- $K_s$  threshold would impart an undesirable selection bias against blue galaxies. To minimize color biases, the J- $K_s$  attribute can be combined with the radial shape attributes to form a new powerful discriminant. First, the color-color plots suggest a more optimum method to use JHK $_s$  colors to measure the “redness” of a galaxy. Galaxies are not only preferentially redder than 0.9 in J- $K_s$ , but they also have H- $K_s$  values,  $>0.2$ , redder than most stars. Hence, we define a “color score” as the color distance in J-H vs. H- $K_s$  space from the line corresponding to J- $K_s = 0.9$  to within a scaling factor. For objects redder than 0.3 in H- $K_s$ , we also factor in the H- $K_s$  color to exploit this feature in the JHK color space. Mathematically, we express the “color score” as:

$$\text{Color score} = [(J - K_s) - 0.9] + \{ [(H - K_s) \geq 0.3] \times [(H - K_s) - 0.3] \}$$

which adds the color “distance” (to within a scaling factor) from the dotted line in Fig. 20. For sources with  $(H - K_s) > 0.3$ , the color score reduces to:

$$\text{Color score} [(H - K_s) > 0.3] = (J - K_s) + (H - K_s) - 1.2$$

Fig. 18 demonstrates the combination of color score and WSH. This combination parameter alone is capable of providing better than 95% reliability ( $K_s < 13.5$ ) with only a few percent loss of galaxies to the total population. We can do better still by using all of the attributes with a decision tree classifier.

#### 4.4 Oblique Decision Tree Classifier

Three classes of attributes have been discussed thus far: radial extent or shape (SH, R1, R23), symmetry or azimuthal shape (WSH, MSH, flux ratios) and flux or photometrics (VINT, “color score”, total flux, and central surface brightness relative to the total flux). To determine the best combination of parameters to use for galaxy discrimination we have a nine-dimensional space to probe. Complicating matters, several of the attributes are highly correlated (e.g., WSH and MSH) and others weakly correlated (e.g., WSH and the bi-symmetric flux ratio), which ultimately prevents simple or

weighted combination of the attributes to form a “super” attribute. We may either combine a few of the attributes that are not strongly correlated (e.g., color score and WSH and R23), e.g., Fig. 18, or employ a decision tree induction method (Breiman et al. 1984) to effectively combine all of the attributes.

In the last few years, decision trees and their close cousins, machine-learning artificial neural networks, have been used by astronomers to aid image classification (e.g., Weir et al, 1995; Odewahn et al. 1992; Salzberg et al. 1995; White 1997). With fast computer technology these methods provide an efficient means to analyze multi-dimensional data. We have adopted one particular type of decision tree, called the oblique-axis decision tree, but there are many others that would also be effective.

Decision tree methods, like “neural networks,” require a training set of pre-classified data composed of all combinations of stars (isolated, double, triple, etc.), galaxies, and artifacts. This “truth” set is used to generate the decision tree, or a structured set of classification rules. Using the analogy of a tree, the rule structure contains “nodes” of branching test points with the final nodes in the tree representing the “leaves” or final classification. For example, one node might represent a test of the WSH score, comparing the score to some threshold, T,

WSH score > T ?

NO: classify as non-galaxy

YES: continue to next node

This is an example of an “axis-parallel” decision. That is to say, the parameter or object attribute embodies a set of hyperplanes (in the multi-dimension phase space) that are parallel to each other. Fig. 19 demonstrates a two-featured, hyperplane: WSH score vs. J mag with galaxies denoted by filled circles and non-galaxies by crosses. The non-galaxies are mostly double stars in this example. The dashed parallel lines represent the axis-parallel “rules.” To the right (or above) the lines are the galaxies; to the left (and/or below) the lines are the false galaxies or non-galaxies. Axis-parallel rules have the advantage of being simple to apply and track within a large complicated tree. But it is obvious from the example plot that a better rule is to use an “oblique” line separating the two populations or features. The solid line in Fig. 19 is an example of an oblique-axis

ruling. An oblique decision tree uses both axis-parallel and oblique-axis tests at the nodes. Mathematically, the node test has the form:

$$\sum_{j=1}^n (a_j O_j) + a_{n+1} > 0$$

where object  $O$  possesses  $n$  attributes, with  $a$  coefficients or weights defining the  $n$ -dimensional hyperplane. For the reduced axis-parallel case, the linear sum reduces to  $a_j O_j > T$ . Although oblique hyperplanes are just a series of linear combinations, the total possible number of solutions is very large and thus finding the "correct" one is daunting, if not impossible under some conditions. In fact, the problem is NP-Complete, or ultimately limited by the runtime of the machine. Fortunately, in practice reasonable decision trees can be generated with clever deduction algorithms and techniques to avoid "traps" or local minimum solutions.

One such package was developed by Murthy et al (1994) called OC1, or Oblique Classifier 1. OC1 uses random perturbations to walk around traps and arrive at satisfactory hyperplane solutions for each node. The resultant tree may require "pruning" or stripping of branches that add little to the final classification, or worse, detract from the correct solution due to over-fitting of the training set. OC1 applies pruning methods, e.g., Cost Complexity pruning (Breiman et al 1984), which effectively prunes the decision tree by removing the insignificant or "weak" branches. For the problem of over-fitting, in addition to pruning, the best solution is to minimize the total number of attributes per node. For 2MASS galaxies, nine attributes including the integrated flux characterize each source. The attributes are correlated to various degrees so it is not obvious which can be eliminated from the decision tree process. A principal component analysis does indicate which parameters are key to the success of the decision tree. Additional trial and error experimentation with the training sets provide further clues as to the level of pruning that our decision tree requires. One disadvantage that decision trees have with classification of galaxies is that the final classification does not have an associated uncertainty or probability of being a galaxy. For 2MASS galaxies, we can "assign" a probability by using a weighted average of the decision tree classifications for each band (see below).

The 2MASS star-galaxy separation problem is well suited to an oblique decision tree technique. Accordingly, we have applied the OC1 technique to large data (training) sets of 2MASS extended sources and non-galaxies (stars, double stars, triples, etc.). The training sets are divided into three density domains: low stellar density fields ( $<10^{3.1}$  stars per deg<sup>2</sup> brighter than 14<sup>th</sup> at  $K_s$ ), moderate ( $10^{3.1}$  to  $10^{3.6}$  stars per deg<sup>2</sup>), and high ( $>10^{3.6}$  stars per deg<sup>2</sup> brighter than 14<sup>th</sup> at  $K_s$ ). These are further divided into subsets depending on the integrated flux of the source. The latter step minimizes the severe dynamic range (in flux) that 2MASS must consider, from the brightest galaxies ( $K_s < 9$ ) to the faintest galaxies ( $K_s > 14$ ). The training sets are large and diverse (the low-density set contains over 15,000 objects, covering some 280 sq. degrees) and thus provide a suitable induction test bed for the decision tree algorithm. We find that the OC1 decision tree classifier improves the galaxy catalog reliability by several percent compared to just using simple CART or axis-parallel tests. The trend persists in regions of high stellar number density where double and triple stars become a major contaminant. Future work to refine the decision trees will focus upon further pruning of the trees and upon possible elimination of “weak” and highly correlated attributes. It may also prove fruitful to evaluate other decision tree methods (for example those developed by Weir et al. 1995; Fayyad 1994) and, possibly, neural network methods, particularly if morphological classification is attempted with 2MASS imaging data.

## 5. 2MASS Extended Source Objects

The 2MASS extended source database is predominantly composed of galaxies and, at the 1 to 20% level depending on the stellar number density, double and triple stars. Large-angular size Galactic objects, such as HII regions, stars with nebulosity, planetary nebulae, reflection nebulae, etc., are relatively rare and generally confined to the Galactic plane and a few other star formation sites.

The final component to the extended source catalog is a class of objects that are collectively referred to as “artifacts.” These false signals generated by bright stars, transient phenomenon, such as meteor streaks, and upper atmospheric “clouds” (or “airglow” in the infrared) which occur at all spatial frequencies. Most artifacts associated with bright stars are easily identified within the 2MASS database using simple geometric

rules, but the occasional false detection slips in. Meteor streaks are more difficult to identify using automated rules, but in general their frequency is low. Airglow not only generates false detections (especially under severe conditions), but it also significantly affects the photometry of real sources. Overall, artifacts generally add less than 1 - 5% to the total contamination under normal conditions. Examples of 2MASS galaxies and various kinds of artifacts are given below.

## 5.1 Galaxies

The 2MASS extended source database contains galaxies ranging in brightness from  $K_s = 9^{\text{th}}$  to  $14^{\text{th}}$  mag. This flux range is constrained by the sensitivity of the survey and size limitations fixed by the scale of the Atlas Image and the background removal process. Galaxies as large as 3 or 4' in diameter may be processed (if they are predominantly located near the center of the coadd) but only the inner 2' or so is examined in detail. For very large galaxies ( $>5'$  in size), such as the Messier objects, the only processing is to extract pieces of coadd images associated with the galaxy. At the faint end of the flux spectrum, the size of the galaxies is limited by the resolution of the survey,  $\sim 2'$ , and the elongation or asymmetry in the PSF.

In Fig. 20-22 a representative sample of galaxies from low stellar number density fields is shown with their  $K_s$ -band postage stamp images. The data come from scans passing through the Abell 3558, Hercules, & Abell 2065 clusters, as well as random non-cluster fields. A wide range in morphology, surface brightness and integrated flux comprise the sample. Fig. 20 shows the brightest galaxies, ranging in total K-band flux from  $9^{\text{th}}$  to  $13^{\text{th}}$  mag. Each image is  $60'' \times 60''$ , and we show examples of several morphological classes: elliptical (E), lenticular (S0, SA0), spiral (S), and complex, including double nucleus, interacting and pre-merger systems. The next set of galaxies, Fig. 21, represent the faint limit at which the extended source catalog is both reliable ( $>98\%$ ) and complete ( $>90\%$ ), with  $K_s$  mags ranging from  $13^{\text{th}}$  to  $13.5^{\text{th}}$  mag. The size of each image is  $30'' \times 30''$ . The final set of low density galaxies (Fig. 22) represent the faintest galaxies resolved with 2MASS, with  $K_s$  mags ranging from  $13.5^{\text{th}}$  to  $15^{\text{th}}$  mag, corresponding to a SNR range between 4 and 8. Each image is  $20'' \times 20''$  in width. The lowest surface brightness galaxies belong to this set, which are generally detected only in



J-band due to the blue color of most LSB-type. For example, the last four galaxies in the set are the result of J-band detections.

When the source density is high, the confusion noise approaches the level of the atmospheric thermal background noise (see Appendix C). The probability of triple or multiple stars is significant and the ability to distinguish galaxies from multiple groupings of stars is limited. Nevertheless, a reliability of >80% is possible for most of the Galactic plane. Fig. 23 gives examples of galaxies found in the Galactic plane, and for comparison, false extended sources (e.g., triple stars) found in the same areas. For the upper panels, the approximate Galactic coordinates are  $(240^\circ, +4.5^\circ)$ , corresponding to a density of 4500 stars  $\text{deg}^{-2}$  brighter than 14<sup>th</sup> mag, and a differential confusion noise equivalent of 0.7 mag in a 10'' aperture (see Appendix C). The integrated  $K_s$ -band fluxes range from 11.8 to 13.8 mag. The estimated 2.2  $\mu\text{m}$  extinction is 0.07 mag and the J- $K_s$  reddening is  $\sim 0.15$  mag. Closer to the Galactic center (Fig. 23 middle panels), coordinates  $(12^\circ, +5.0^\circ)$ , the density of stars is over 30,000 per  $\text{deg}^2$ , resulting in an equivalent differential K-band confusion noise of nearly 2 mags—yet galaxies are still detected by 2MASS. The estimated 2.2  $\mu\text{m}$  extinction is now 0.2 mag and the J- $K_s$  reddening is 0.38 mag. Note the significant stellar contamination to the local environment of the galaxies. The integrated  $K_s$ -band flux ranges from 11.0 to 12.8 mag, indicative of confusion noise limits on the faint end discovery spectrum. False detections are dominated by multiple stars (mostly triples and quadruples), a representative set is shown in the lower panels, Fig. 23.

## 5.2 Galactic Extended Sources

Nebulosity associated with bright stars (e.g., H II regions, PNs) and with molecular clouds (reflection nebulae, YSOs) typically appear as very bright and large extended sources (Fig. 24). Since these objects are primarily located deep in the Galactic plane, confusion with stars is significant.

## 5.3 Bright Stars and Artifacts

Bright stars are historically a major nuisance to any image-based survey. Off-axis stray light can land just about anywhere on the focal plane, while dense concentrations of

light (e.g., diffraction spikes) are distributed geometrically with respect to the optical axis. Features referred to as “glints” and “ghosts” are focused or semi-focused reflections of light that appear as slightly asymmetric point sources or flattened (low surface brightness) extended sources. Not only do bright 2MASS stars ( $K_s < 9^{\text{th}}$ ) produce diffraction spikes, halos, glints and ghosts, but for the brightest stars ( $K_s < 5^{\text{th}}$  mag, which is approximately the saturation limit of the 2MASS survey), they generate horizontal stripes that persist along the entire cross-scan (east-west axis) of the scan, or a total of  $8.5'$  in length. Finally, bright stars induce another feature unique to infrared arrays: latent residual or persistence ghosts. The central core of a bright star leaves a residual signal after the array has been read out. The residual persists for several seconds (and for the brightest stars, many tens of seconds). What this means for 2MASS survey data is that a bright star will leave a “trail” of persistence ghosts as the telescope shifts in declination. All of these bright star artifacts, many of which look like perfectly reasonable galaxies, must be removed within the pipeline reduction process to minimize false detections. Halos, stripes and spikes have a well-determined geometry with respect to their progenitor. GALWORKS determines their extent by measuring their surface brightness and using limits based on the estimated total flux of the star and the expected confusion noise as traced by the stellar number density.

Diffraction spikes extend several arcminutes for very bright stars; see for example Fig. 25, which shows a  $4^{\text{th}}$  magnitude star in a J-band Atlas image. Note the three horizontal stripes extended and flaring across the entire  $8.5'$  of the field. Also note the persistence ghosts trailing to the south of the bright star. An even more dramatic example of spikes, ghosts, halo and stripes is seen in Fig. 26, which shows two adjacent J-band coadds with a  $K_s = -1$  mag star ( $\beta$  Pegasus) straddling the boundary. The vertical spikes extend well beyond the coadd boundaries, while the halo emission completely dominates both coadds. The persistence ghosts (trailing to the south of  $\beta$  Peg) appear nearly as bright as field stars. The influence of  $\beta$  Peg extends across scan boundaries as well. Internal telescope reflections produce stripe features extending over  $1^\circ$  in radius from the center of  $\beta$  Peg (see Fig. 27), presenting a major clean-up challenge for artifact removal. It may be that in the vicinity of the brightest stars ( $K_s < 0$  mag) in the infrared sky, the

reliability will inevitably suffer. Fortunately, there are only a handful of these grisly stars.

Many meteor streaks have the unfortunate property of high surface brightness coupled with severe elongation—similar to large highly inclined spiral galaxies. Fig. 27 demonstrates transient streaks in two different J-band coadds. Note the sharp boundaries for the bright streak and the episodic flaring for the fainter streak. The latter is, in fact, associated with  $\beta$  Pegasus (Fig. 26), discussed above. The one critical difference that meteor streaks have with real edge-on spiral galaxies is that usually multiple detections (in some cases several hundred sources) occur along the streak which can in principle be identified with simple database queries and cleaned from the catalogs accordingly.

Additional artifacts are produced by bright stars, meteor streaks, background gradients (for example, airglow “bumps” that are not removed) and bright to moderately bright stars on the edges of scans. Fig. 28 illustrates some of the kinds of artifacts found in the extended source database. The first two (reading left to right) are the result of a “ghost” or “glint”, most prominent in J band, to the southwest of the 8<sup>th</sup> – 9<sup>th</sup> mag progenitor star. The third column shows a false detection due to a flared diffraction spike from a star on the edge of coadd. The 4<sup>th</sup> and 5<sup>th</sup> columns are examples of faint stars or faint galaxies located on or within the boundary of a horizontal stripe or meteor streak. The final column is a faint star boosted up by background airglow (note the prominent H-band emission). The airglow artifact is probably the most insidious class of false detection since it is so difficult to discriminate from real galaxies or real interstellar nebulosity. The only way to minimize their effect is to avoid data with significant airglow. Because the airglow at H-band shows the most severe variations, we therefore exclude sources identified in the H-band only.

### Acknowledgements

This publication makes use of data products from 2MASS, which is a joint project of the Univ. of Massachusetts and the Infrared Processing and Analysis Center, funded by the NASA and the NSF. This work was supported in part by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with NASA.

## Appendix A. Galaxy Photometry Error Tree

The fundamental limits to 2MASS galaxy photometry are set by the accuracy to which the background level can be determined, by the total signal to noise in the aperture used to report the flux of a galaxy, and by the zero-point calibration used to adjust raw magnitudes onto a standard scale. Errors in the background removal process are the principal cause of errors in the photometry. In the near-infrared, the background levels generally fluctuate due to airglow emission, sometimes at very high spatial frequencies, particularly at H-band. In addition to airglow, 2MASS images include electronic (“pickup”) noise that is correlated and can, at times, be comparable to the airglow component. Electronic pickup refers to image features that are not of astronomical origin, but instead are associated with the camera electronics. These features may manifest as periodic horizontal stripes or abrupt jumps in the background level. The impact is mostly felt in the galaxy photometry, where a systematic bias is introduced. This bias can be as large as 20% in some cases, but is more typically 3 to 7% (usually an overestimate of the galaxy flux). Examples of both forms of background enhancement are given below.

### A.1 Summary of Photometric Error

The measurement error in summing up the flux over  $n$  pixels in the coadd is

$$\sigma_{pix} \times \sqrt{1.7 \times 4 \times n}$$

where  $\sigma_{pix}$  is the measured individual pixel noise for the usual case where the galaxy flux is negligible compared to the background flux. The factor of 1.7 corrects for the smoothing introduced in the coadd by the frame resampling and construction process, and the factor of 4 results from the correlation of the flux in 1” coadd pixels, due to the 2” camera (raw frame).

GALWORKS also models and subtracts variations in the background using a polynomial fitting scheme described in §3.1. If the background variations are small on time scales of a few seconds, the fitting procedure is limited primarily by  $\sigma_{pix}$ . In practice the fitting procedure can match the background over regions as small as  $\sim 3'$ . Empirical tests indicate that the fit is accurate to about  $0.03 \sigma_{pix}$  over these size scales. Since errors in the background contribute systematically to all pixels in a galaxy smaller than  $3'$ , this

error contributes in proportion to the number of pixels  $n$ . The total photometric noise is the sum of these two terms

$$\sigma_{pix} \times [(0.03 \times n) + (3.4 \times \sqrt{n})]$$

Consider for example a K~13th mag galaxy. The typical size of this galaxy is about 15'' in diameter, and the typical background pixel noise is ~1 dn (10 electrons) in high Galactic latitude fields, giving a resultant signal to noise ratio of ~12, or a photometric error of 8%.

If the background variation requires a higher-order polynomial, and is well-fit by a cubic (3<sup>rd</sup> order), which is normally the case except for H-band observations with significant airglow emission, then the coadd is effectively further divided into 3 or 4 sub-regions with the resulting errors then being ~3-4 times higher. The resulting noise in the determination of the background is then  $\sim 0.03 \times \sigma_{pix}$ .

If the background variation requires a higher-power polynomial than a cubic, then the error in the background will no longer be determined by the pixel noise. Instead, the background error will result from the residual error after a cubic polynomial is fit. This will be the case for severe airglow variation, which can vary on high spatial frequency scales (see Fig. A.1), and from correlated "electronic" noise (see Fig A.1).

In order to keep the noise in determining the background level from increasing the total photometric noise by more than 10%, that noise must be less than half the Poisson background noise component. The noise in determining the background level must then be less than about ~0.05 DN, using worst case limits (but still acceptable for the survey as a whole) for the aperture size and background levels. Thus, the typical noise in determining the background is generally less than this limit, except when the airglow is severe or significant correlated noise is present.

## A.2 Adaptive Aperture Errors

Adaptive apertures (i.e., Kron and isophotal) come in two different shapes: circular and elliptical. Circular apertures are fully described by the radius, fit to the desired isophote. Elliptical apertures are described by three parameters, radius (semi-major axis), axis ratio and position angle, fit to the desired isophote. The fitting procedures introduce additional uncertainties.

For the circular case, the radius has an error that is driven by background noise and contamination (i.e., stars near or within the isophote). Errors in the determination of that radius dominate the photometric error tree. Consider the best case of a galaxy with an exponential profile. The integral flux over a circular aperture is

$$C \times [1 - (1+r)e^{-r}]$$

where  $r$  is the aperture radius divided by the radial scale length of the galaxy profile. The flux error caused by the error in radius determination is

$$df / f = \frac{r}{[e^r - (1+r)]} dr$$

For a disk scale length of  $2''$ , and typical isophotal 20 mag per sq.arcsec aperture radii of  $6-8''$ , the flux error ranges between

$$df / f \approx 0.15 - 0.05 dr$$

The width of the circular isophote,  $dr$ , is  $\sim 1''$ , so the radius determination gives a flux error of 5-15%. The Poisson photon noise in an aperture of this size gives typical flux errors of  $\sim 5-7\%$ , and hence the radius uncertainty dominates the photometric error for this (typical  $K \sim 13$  mag) case.

For elliptical aperture photometry, in addition to the radial uncertainty, two additional parameters (axial ratio and position angle) add to the total error budget. However, there is some positive compensation due to the fact that the orientation of galaxies is elliptically shaped; thus, optimal elliptical apertures minimize the background noise contribution and systematics due to measurement error in the background itself (§A.1 above). Analysis of the 2MASS extended source database reveals that for 2MASS galaxies brighter than  $\sim 13^{\text{th}}$  mag, the elliptical aperture photometry gives the most precise measurement of the galaxy flux. Note however that the elliptical aperture is vulnerable to contamination, especially in high source density regions, which makes it susceptible to large uncertainties if the ellipse parameters are inaccurate. For faint galaxies, the PSF shape and image resolution generally favor a circular aperture to that of an ellipse because the galaxies are so small that the ellipse parameters are highly uncertain.

## Appendix B. Extended Source Catalog Parameters

The 2MASS Extended Source database contains over 400 parameters per source, most of which are related to photometry. The public-accessible catalogs are derived from this database. In Table 1 a condensed list is given of the most important parameters in the catalogs. The list includes positions, ellipse-shape geometry, photometry, surface brightness, star-galaxy discrimination scores, symmetry, and probability of extendedness. The extended source probabilities are derived from the decision trees (see §4.4). The parameter name represents the actual designation used in the catalogs and database.

For galaxy photometry and colors, we recommend the fiducial  $K_s$ -band 20 mag/sq.arcsec ellipse-aperture measurements only for brighter galaxies. For smaller galaxies (comprising most of the catalog) the fixed  $R=7''$  circular-aperture values are the most robust flux measures; see Table 1. For each flux measure there is an associated confusion flag, where a non-zero value indicates stellar contamination.

## Appendix C. Stellar Number Density and Confusion Noise

As the surface density of stars exponentially increases near the disk of the Galaxy, the probability of source contamination increases accordingly. Likewise, near the plane, the "sky" background is elevated in surface brightness due to faint undetected stars driving up the overall sky brightness and consequently the mean noise amplitude. Both surface density effects limit the source detection or completeness, and overall reliability. A convenient gauge for the severity of contamination or level of source confusion, is the "confusion noise," and the associated change in surface brightness relative to the pole, expressed in magnitude units. As the confusion noise becomes appreciable, it is one of the primary limits on galaxy detection. Moreover, confusion decreases the accuracy of both flux and position estimation. It is therefore important to understand the confusion noise in terms of the ability to detect isolated sources and in terms of identification of real extended sources, both of which require threshold adjustment with confusion noise level.

### C.1. Estimation of Confusion Noise

To estimate the additional component of “confusion” noise to the total, we adopt the methodology of Hacking et al. (1987). The idea is to integrate the expected number of sources (with some flux distribution) within the 2MASS effective beam,  $\Omega_e$ . We may approximate the stellar flux distribution with a power law of index  $\alpha$ ,

$$N(f) \propto f^{-\alpha}$$

where  $N$  is the integrated number density at flux  $f$ . The aggregate variance due to background sources in the beam, derived from  $\delta f$ ,  $d\Omega$  and the differential stellar number density, is then

$$\sigma_{conf}^2 \propto \frac{\alpha \Omega_e (q\sigma)^{2-\alpha}}{2-\alpha}$$

where  $q$  is the  $n$ - $\sigma$  cutoff at which the distribution outliers are excluded from the noise calculation (typically  $3 - 5\sigma$ ). The source density index  $\alpha$  is approximately equal to unity (more precisely  $\geq 0.85$ ) as derived from the log-log slope of the  $N$  vs.  $K_s$  cumulative starcount curve,  $\sim 0.35$  for the NGP and slightly steeper at higher densities (Jarrett 1992). The confusion noise reduces to

$$\sigma_{conf}(\alpha = 1) \propto \Omega_e \times q$$

The change in surface brightness due to the confusion noise is given by  $\sigma_{conf}$  per beam  $\Omega$ , expressed in mag units, or  $\Delta mag$ , where “minimum” confusion noise corresponds to the Galactic pole. The K-band confusion noise as a function of the total integrated stellar number density ( $K < 14$  mag) is plotted in Fig. C.1, described below, assuming  $\alpha = 1$  and  $q = 5$  (i.e.,  $5\sigma$  cutoff of outliers) and a beam size of  $5''$ .

## C.2. Confusion Noise, Stellar Density and Galactic Coordinates

For relatively moderate flux ranges (e.g.,  $V < 18$ ;  $K < 14$ ) basic three-component models of the Galactic stellar distribution adequately describe the number density of dwarf and giant stars comprising “disk” and “spheroid” populations (Elias 1978; Bahcall & Soneira 1980; Garwood & Jones 1987). Here we employ a near-infrared-modified variation on the Bahcall & Soneira model described in Jarrett (1992), which predicts the stellar number density with  $\sim 90\%$  accuracy for most of the sky ( $|b| > 30$ ) and  $K < 14$  mag, and performs adequately ( $\sim 80\%$ ) for the Galactic plane where patchy extinction



ultimately limits the utility of these simple exponential models. The star-count model predicts the stellar number density as a function of the Galactic coordinates, which can then be used to calculate the approximate confusion noise.

A plot of the stellar number density as a function of the Galactic latitude along two separate longitudes ( $50^\circ$  and  $130^\circ$ ) is shown in Fig. C.1. The vertical dotted lines represent the thresholds for what is considered low stellar number density ( $<10^{3.1}$  stars per  $\text{deg}^2$ ), moderate density ( $<10^{3.6}$  stars per  $\text{deg}^2$ ), and high density ( $>10^{3.6}$  stars per  $\text{deg}^2$ ). The limit on high density is partly driven by the relative density of triple stars versus double or single stars. As triple+ stars become appreciable, the ability to distinguish real galaxies from close groupings of stars is greatly diminished. Finally, the confusion noise ( $\Delta\text{mag}$ ) corresponding to the stellar number density is plotted in Fig. C.1, with the confusion noise axis located at the right of the plot.

## References

- Bahcall, J.N. and Soneira, R.M. 1980, Ap. J. Suppl., 44, 73.
- Beichman, C.A., Chester, T.J., Skrutskie, M., Low, F.J., & Gillett, F. 1998, PASP, 110, 480.
- Bertin, E. & Arnouts, S. 1996, A&AS, 117, 393.
- Breiman, L., Freidman, J., Olshen, R., & Stone, C. 1984, Classification and Regression Trees (Wadsworth & Brooks, Monterey, CA).
- Bruzual A., G., & Charlot, S. 1993, ApJ, 405, 538.
- Chester, Jarrett, T.H., Schneider, S.E., Skrutskie, & M., Huchra, J. 1998, BAAS, 30, No. 2, #55.11.
- Cutri, R.M. 1997, Cutri, R.M.1997, in "The Impact of Large Scale Near-IR Sky Surveys, F. Garzon et al. (eds.), Kluwer (Netherlands).
- Elias, J.H. 1978, Ap. J., 223, 859.
- Elias, J.H. 1978, Ap. J., 224, 453.
- Fayyad, U. 1994, in Artificial Intelligence AAAI-94 (MIT Press, Cambridge, MA), 6601.
- Garwood, R. and Jones, T.J. 1987, PASP, 99, 453.
- Hacking, P. 1987, PhD Thesis, Cornell Univ.
- Gardner, J.P. 1998, PASP, 110, 291.
- Gardner, J.P., Cowie, LL, & Wainscoat, R. 1993, ApJ, 415.
- Gardner, J.P., Sharples, R.M., Carrasco, B.E., and Frenk, C.S. 1996, MNRAS, 282, L1.
- Glazebrook, K., Peakcock, J.A., Collins, C.A., & Miller, L. 1994, MNRAS, 266, 65.
- Jarrett, T.H., Chester, T., Cutri, R., Hurt, R., Schneider, S., Hucra, J., 1999, Nature, submitted.
- Jarrett, T.H., 1992, Ph.D. Thesis, University of Massachusetts.
- Jarrett, T.H., Chester, T., Schneider, S. & Huchra, J., in "The Impact of Large Scale Near-IR Sky Surveys, F. Garzon et al. (eds.), Kluwer (Netherlands), 213.
- Jarrett, T. 1998, in The Impact of Near-Infrared Sky Surveys on Galactic and Extragalactic Astronomy", ed. N. Epchtein, (Netherlands: Kluwar), 239.
- Kleinmann, S., et al. 1994, A&SS, 217, 11.

- Kleinmann, S., et al. 1994, in *Infrared Astronomy with Arrays: The Next Generation*, ed. I. McClean (Dordrecht: Kluwar), 219.
- Koo, D.C. 1986, *ApJ*, 311, 651.
- Kron, R.G. 1980, *ApJS*, 43, 305.
- McLeod, B. A., & Rieke, M.J. 1995, *ApJ*, 454, 611.
- Murthy, S.K., Kasif, S., & Salzberg, S. 1994, "A System for Induction of Oblique Decision Trees", *JAIR*, 2, 1.
- Odewahn, S.C., Stockwell, E.B., Pennington, R.L., Humphreys, R.M., & Zumach, W.A. 1992, *AJ*, 103, 318.
- Ramsay, S.K., Mountain, C.M., & Geballe, T.R. 1992, *MNRAS*, 259, 751.
- Schneider, S.E., Huchra, J.P., Jarrett, T.H. and Chester, T.J. 1997, in *The Impact of Large Scale Near-IR Sky Surveys*, F. Garzon et al. (eds.), (Dordrecht: Kluwar), 187.
- Skrutskie, M., et al. 1997, in *The Impact of Large-Scale Near-IR Sky Survey*, ed. F. Garzon et al. (Dordrecht: Kluwar), 25.
- Salzberg, S., Chandar, R., Ford, H., Murthy, S., & White, R. 1995, *PASP*, 107, 279.
- Stetson, P.B. 1990, *PASP*, 102, 932.
- Valdes, F. 1982, *SPIE Proc. On Instrumentation in Astronomy IV*, 331, 465.
- Weir, N., Fayyad, U.M., & Djorgovski, S. 1995, *AJ*, 109, 2401.
- White, R.L. 1997, in *Statistical Challenges in Modern Astronomy II*, ed. Babu, G. & Feigelson, E. (New York: Springer), 135.

## Figure Captions

Fig. 1.—Stellar SH ridgeline for a scan passing through the Hercules galaxy cluster. The scan coordinate corresponds to the declination axis, or effectively, the time axis. The SH or  $(\alpha \times \beta)$ , in the generalized exponential function; see text for details, per source is shown for the  $K_s$  (top), H (middle) and J (bottom) bands. Mean SH values and their associated uncertainties (statistical rms of distribution) are denoted with filled circles and error bars. The stellar ridgeline is defined by the mean SH values. Resolved Hercules cluster galaxies, which have intrinsically large SH values, are the points that scatter above the ridge line.

Fig. 2.—Stellar SH ridgeline for a 2MASS scan with poor atmospheric “seeing.” The scan coordinate corresponds to the declination and time axis. The SH per source is shown for the  $K_s$  (top), H (middle) and J (bottom) bands. Mean SH values and their associated uncertainties (rms of distribution) are denoted with filled circles and error bars. See Fig. 1 for further details.

Fig. 3.—2MASS Extended Source Processor I/O flowchart.

Fig. 4.—Detailed flowchart for the 2MASS Extended Source Processor, GALWORKS.

Fig. 5.—2MASS Atlas Image decomposition schematic for background fitting. The J, H,  $K_s$  raw images have  $512 \times 1024$  pixels ( $\sim 8.5 \times 16'$ ) each. The first step is to resample the image with an  $8 \times 8$  median filter. A cubic polynomial is then fit to the surface defined by dividing the filtered image into three chunks: upper, middle and lower sections, with 50% overlap between the middle and upper segments, and middle and lower segments. The final background solution results from a weighted-average (overlap dependent) stitching between the three segments.

Fig. 6.—Example of a raw J, H and  $K_s$ -band images, their corresponding background solutions and residual (background subtracted) images. The gray-scale stretch ranges

from  $-2\sigma$  to  $5\sigma$  (noise scaling) of the median background level. Notice the prominent “airglow” background gradients in the raw H-band image (middle panel) and low-level-high-frequency ridges in the residual image. The background level of the H and  $K_s$  images are approximately  $4\times$  larger than that of J-band due to thermal emission from the atmosphere.

Fig. 7.—Bright stars with associated nebulosity. The first five sources come from the Orion trapezium region and the last three from the Large Magellanic Cloud. The upper row shows J-band postage stamp images, middle row the H-band and bottom row the  $K_s$ -band images. Each image is  $50''\times 50''$  in angular size. The integrated  $K_s$  flux for the stars range from  $5^{\text{th}}$  to  $7^{\text{th}}$  mag.

Fig. 8.—Low central surface brightness galaxies. Typical set of galaxies detected and extracted with the LCSB processor. The upper row shows J-band postage stamp images, middle row the H-band and bottom row the  $K_s$ -band images. Each image is  $25''\times 25''$  in angular size. The integrated  $K_s$  flux ranges from 13.8 to 15.2 mag.

Fig. 9.—The expected number of stars (solid line), galaxies (dashed line), double stars and triple stars (dotted lines) with Galactic latitude. The longitude is fixed at 90 degrees. The calculations are based on the star-count models of Jarrett (1992). Double stars, mostly sky-projected associations, represent "primary-secondary" separations of less than  $6''$  and triple stars with less than  $10''$  (the 2MASS PSF for comparison has a FWHM  $\sim 2-3''$ ). The field galaxy counts are based on 2MASS data.

Fig. 10.—Distribution of stars, multiple stars and galaxies in the J-band SH versus magnitude parameter plane. The sources do not come from the same sample: the triple stars are derived from high stellar source density fields in the Galactic plane, while the galaxies come from low confusion areas. For most of the sky, stars generally outnumber galaxies by a ratio of 10:1 for J brighter than  $15^{\text{th}}$  mag.

Fig. 11.—Examples of large Virgo galaxies as seen in the 2MASS J, H and  $K_s$  image data. The color composite is derived in the standard fashion: blue == J band, green == H band, red ==  $K_s$  band. Each image is 100'' in angular size, the maximum size for 2MASS extended source "postage-stamp" images.

Fig. 12.—Comparison of 2MASS double stars and galaxies of comparable brightness. The upper panel shows a variety of doubles encountered in the survey. The lower panel shows galaxies with approximately the same total integrated flux as with the double stars.

Fig. 13.—Cartoon representation of measuring the radial extent of a "double star" using adaptive masking. A wedge-shaped mask, vertex anchored to the primary star and rotated through all angles, is used to selectively block pixel areas. At each rotation angle, the radial profile is reconstructed (excluding pixels from masked area) and is fit with the generalized exponential function (Eq. 1). The WSH score corresponds to the minimum  $SH (\alpha \times \beta)$  of the set defined by the rotation angle, which ideally corresponds to the optimum masking of the secondary star. Hence, the WSH value approaches that of an isolated star, although it is never as small as that of an isolated star since the secondary star contaminates the true centroid position of the primary star (depending on the pair relative brightness and separation).

Fig. 14.—Distribution of multiple stars and galaxies in the J-band WSH score versus magnitude parameter plane. Note: the sources do not come from the same sample: the triple stars are derived from high stellar source density fields in the Galactic plane, while the galaxies from low stellar source density fields.

Fig. 15.—Distribution of multiple stars and galaxies in the J-band R23 score versus magnitude parameter plane. Note: see Fig. 14 notes.

Fig. 16.—Histogram of the J- $K_s$  color distribution for galaxies, double and triple stars in low, moderate and highly confused areas. The low-density data, corresponding to  $<10^{3.1}$  stars  $\text{deg}^{-2}$  brighter  $K = 14^{\text{th}}$  mag, come from a diverse set of fields comprising some 250

deg<sup>2</sup>. The moderate density data,  $10^{3.1} - 10^{3.6}$  stars deg<sup>-2</sup> come from fields comprising some 150 deg<sup>2</sup>. The high density data,  $>10^{3.6}$  stars deg<sup>-2</sup> come from fields comprising ~60 deg<sup>2</sup>.

Fig. 17.—J-H vs. H-K<sub>s</sub> color plane distribution for sources in minimally and highly confused areas. Galaxies are denoted with open circles, double stars with filled triangles and triple stars with crosses. The solid line demarks the main sequence tracks (dwarfs lower track, giants upper track). The K-correction track for spirals follows the solid line and diamond symbols, where the diamonds denote intervals of 0.1 in redshift (z). The reddening vector is shown for the high density color-color distribution (right panel), where dust extinction is expected to be significant.

Fig. 18.—Distribution of stars and galaxies in the “color score + WSH” parameter space (see text). See Fig. 16 for symbol descriptions.

Fig. 19.—An example of a two-featured data hyperplane set that represents a decision tree node. A subsection of the WSH score—J magnitude plane for galaxies (denoted with filled circles) and non-galaxies (such as double stars; denoted with cross symbols) is shown. Axis-parallel planes are represented with dashed lines and the best-fit oblique plane is represented with a solid line.

Fig. 20.—Bright 2MASS galaxies as seen in the K<sub>s</sub>-band. The sequence is arranged in order of integrated K<sub>s</sub>-band flux, ranging from 9<sup>th</sup> to 13<sup>th</sup> mag. Each image is 60'' in angular width.

Fig. 21.— 2MASS galaxies at the K<sub>s</sub>-band sensitivity requirement limit,  $K_s = 13.0$  to  $13.5$  mag. Each image is 30'' in angular width.

Fig. 22.—Faint 2MASS galaxies as seen in the K<sub>s</sub>-band. Each image is 20'' in angular width.

Fig. 23.—2MASS extended sources in the Galactic plane. The upper panel shows galaxies found in the Zone of Avoidance at approximate Galactic coordinates:  $240^\circ$ ,  $+4.5^\circ$ , corresponding to a density of 4500 stars per  $\text{deg}^2$  brighter than  $14^{\text{th}}$  mag. The middle panel shows galaxies found near the Galactic center bulge ( $\sim$ Galactic coordinates:  $12^\circ$ ,  $+5.0^\circ$ ), corresponding to a density of  $>30,000$  stars per  $\text{deg}^2$  brighter than  $14^{\text{th}}$  mag. For comparison, a set of false galaxies—triple stars—are shown in the bottom panel. Each J, H,  $K_s$  3-color composite image is  $50''$  in angular width.

Fig 24.—Galactic extended sources. J, H,  $K_s$  3-color composite images of H II regions (upper panel), stellar clusters and emission nebulosity (middle panel), and reflection nebulae and YSOs (bottom panel).

Fig. 25.—J-band Atlas (coadd) image of a  $4^{\text{th}}$  magnitude star. The image size is  $8.5 \times 16'$ . Features associated with the bright star: halo emission, N-S-E-W diffraction spikes, three horizontal stripes, glints/ghosts, and persistence ghosts (trailing to the south of the star).

Fig. 26.—Consecutive J-band Atlas (coadd) images of  $\beta$  Pegasus, a  $-1$  mag star. The star lands on the in-scan boundary of two coadd images. The total area is approximately  $8.5' \times 25'$ . Note the prominent halo emission, N-S-E-W diffraction spikes, three horizontal stripes, glints/ghosts (particularly to the northeast), and persistence ghosts trailing to the south of the star.

Fig. 27.—Meteor and bright star streaks as seen in the J-band. The images are  $8.5'$  across. The meteor streak is the left panel, while the right shows a streak associated with  $\beta$  Pegasus, a very bright star located nearly one degree away.

Fig. 28.—Example of “artifact” or false extended source detections. The upper panel shows J-band, middle panel H-band and the bottom panel  $K_s$ -band. Each image is  $30''$  in width. The first two columns are the result of a “ghost” or “glint” to the southwest of the



progenitor star; 3<sup>rd</sup> column shows a false detection due to a flared diffraction spike from a star on the edge of coadd; 4<sup>th</sup> and 5<sup>th</sup> columns are examples of real sources located on or within the boundary of a horizontal stripe or meteor streak; last column shows a faint star bathed within background airglow emission.

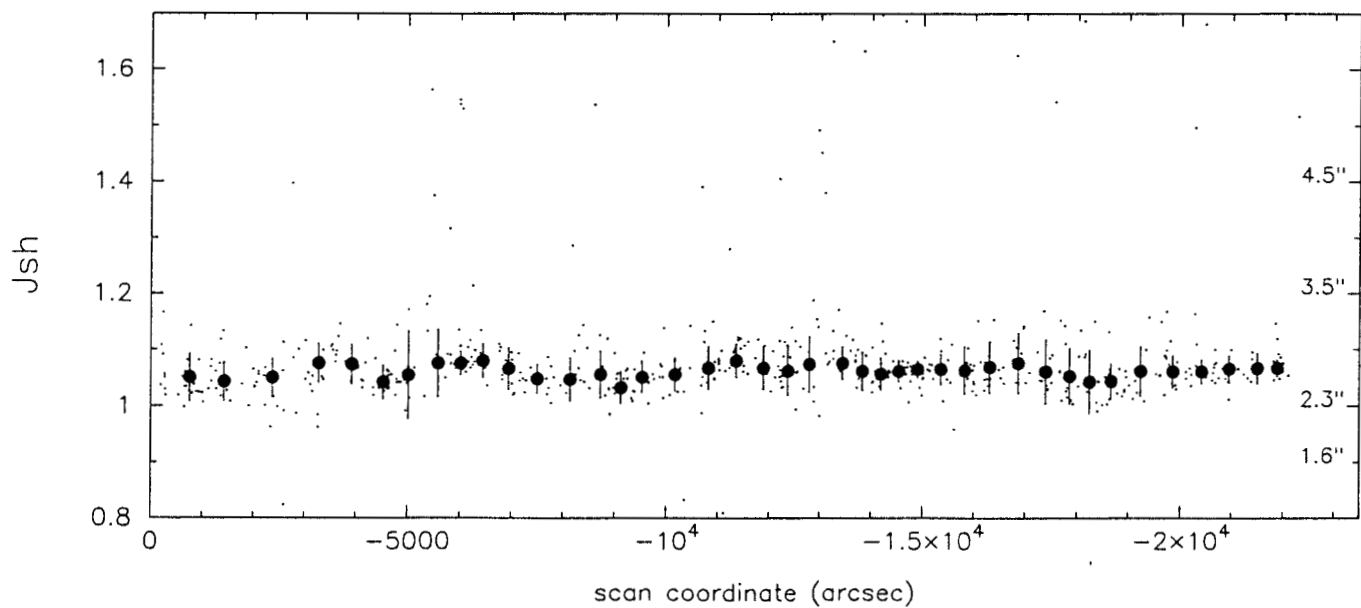
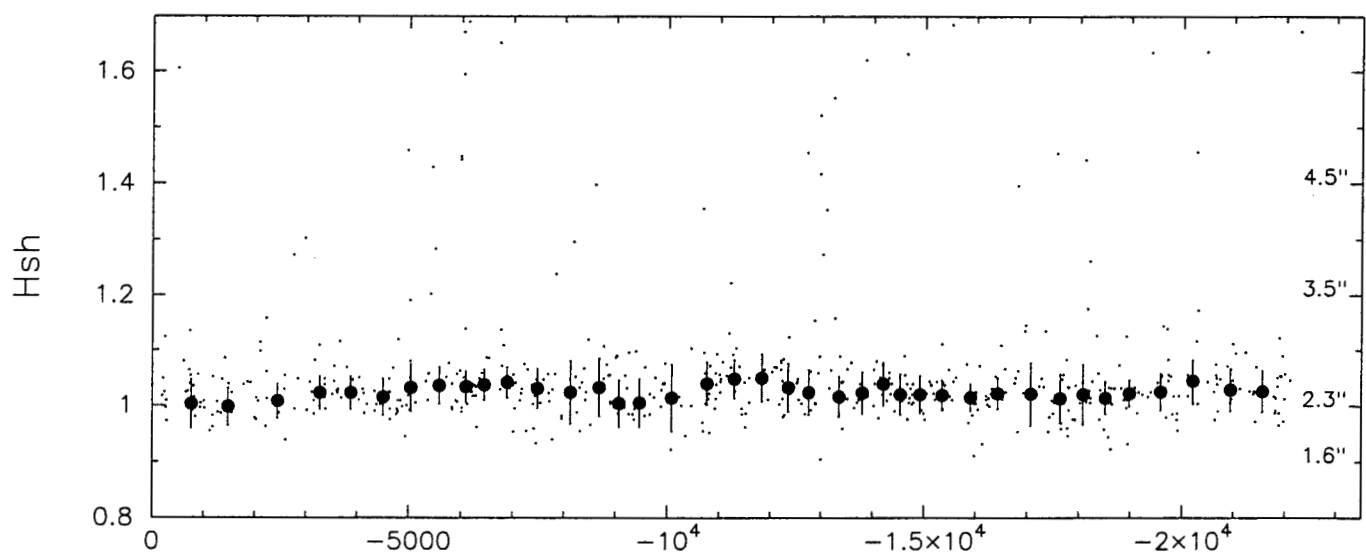
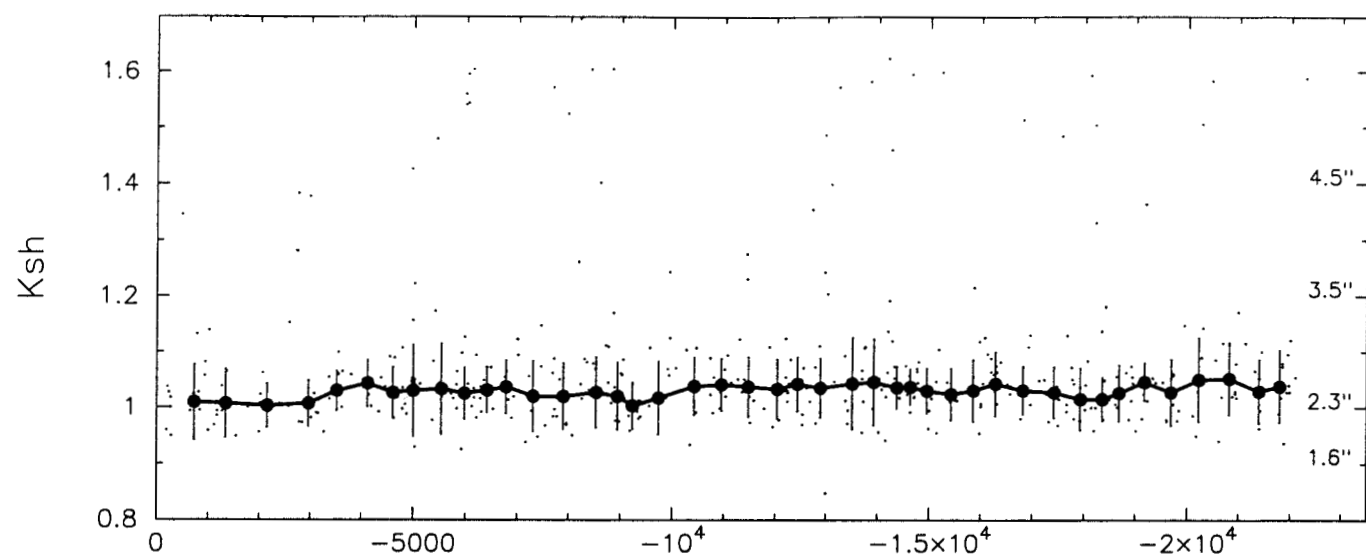
Fig. A.1.—Example of severe background "contamination." The left image shows a case of severe airglow emission as seen in the H-band Atlas image. The airglow fluctuates on scales of  $\sim 1'$ . The right image shows a case of correlated "electronic" (non-astronomically related) noise ridges along the right side of a J-band image. The ridges are the result of one array frame quadrant (NICMOS arrays have four quadrants) having elevated pixel values, which induce sinusoidal waves with drift scanning and coaddition of individual frames (i.e., the drift step size and quadrant size are constructively synchronized). Note also the frame-edge features seen in the right panel.

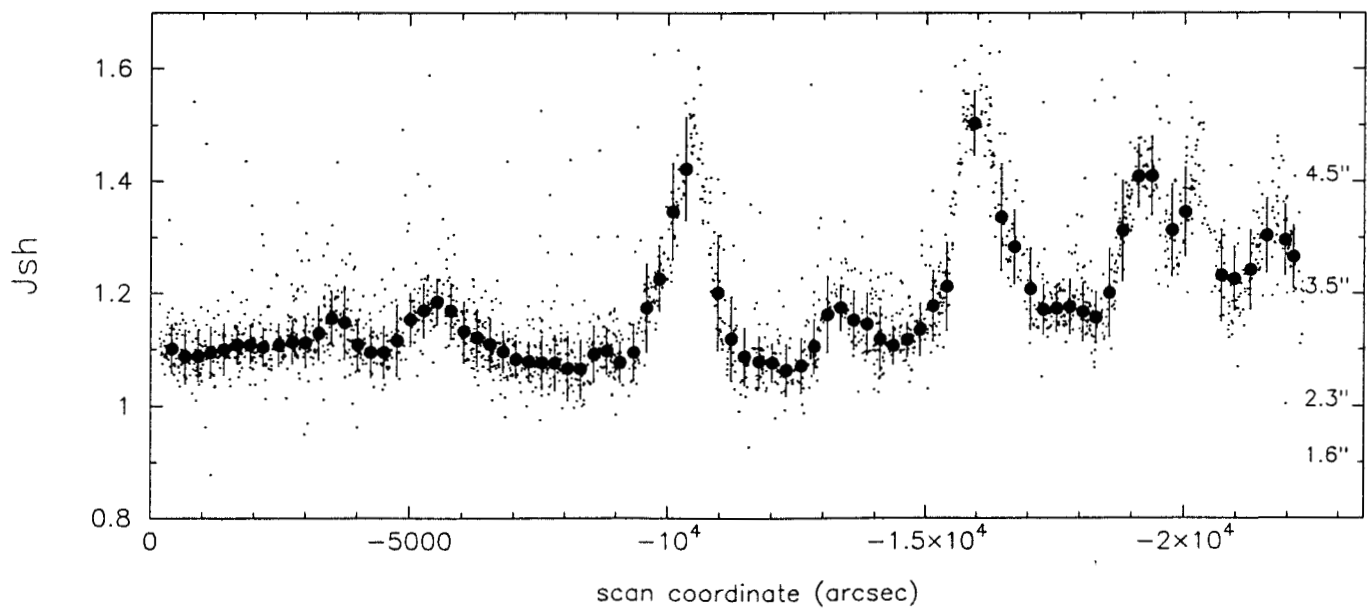
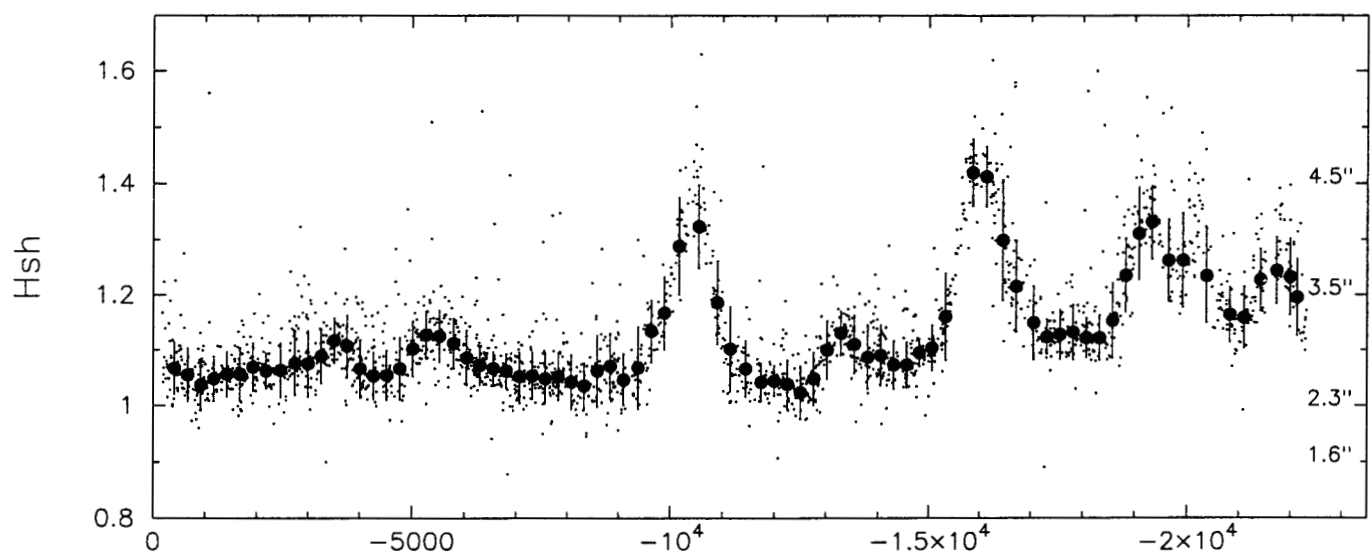
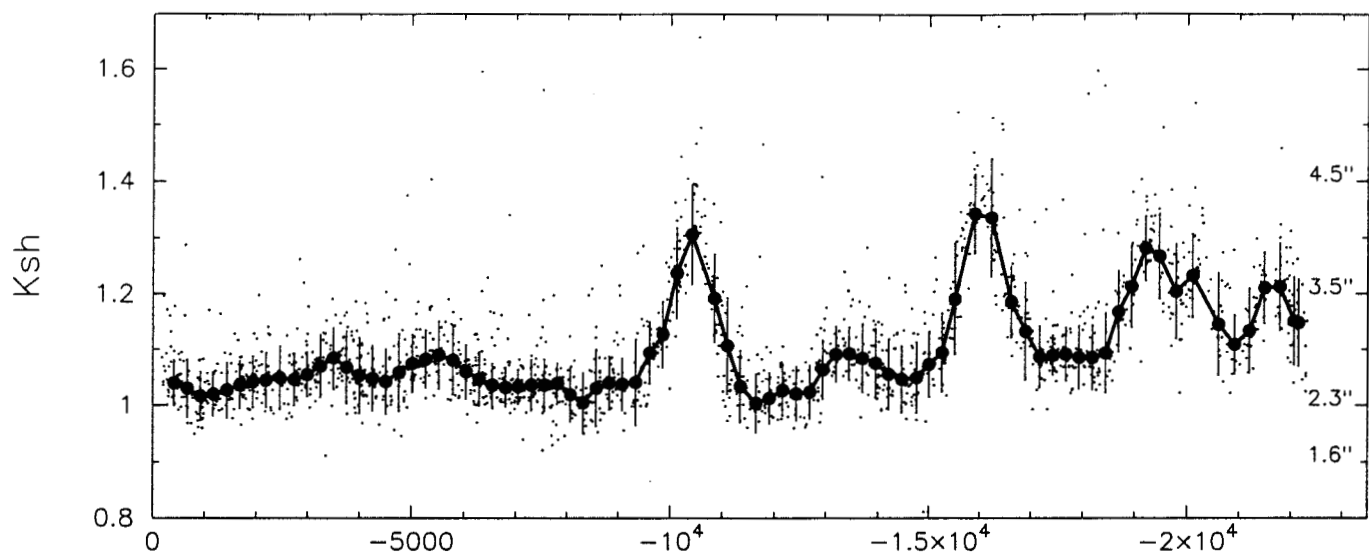
Fig. C.1.— The predicted stellar number density,  $\log [\text{stars deg}^{-2}]$  brighter than 14<sup>th</sup> mag at  $K_s$ , as a function of the Galactic latitude. Two sets of solutions are shown, that of 50° Galactic longitude (solid line) and 130° Galactic longitude (dashed line). The dashed lines define regions of low ( $< 3.1$ ), moderate (3.1 to 3.6) and high ( $> 3.6$ ) stellar number density. The estimated confusion noise ( $\Delta\text{mag}$ ) as a function of the stellar number density is denoted by the dot-dashed line, assuming index  $\alpha = 1$ , outlier cutoff  $q = 5\sigma$ , and beam size of  $5''$ .

TABLE 1. 2MASS Extended Source Catalog Key Parameters	
Parameter	Description
ra	right ascension (J2000) in degrees; based on peak pixel
dec	declination (J2000) in degrees; based on peak pixel
density	Atlas image stellar number density: $\log(\text{stars per deg}^2)$ with $K_s < 14$ mag
<band>_ba	<band> minor/major axis ratio fit to the $3\sigma$ isophote
<band>_phi	<band> angle to $3\sigma$ major axis (E of N)
j_chif_ellf	%chi-fraction for ellipse fit to $3\sigma$ J-band isophote
k_chif_ellf	%chi-fraction for ellipse fit to $3\sigma$ $K_s$ -band isophote
sup_ba	minor/major axis ratio fit to $3\sigma$ super-coadd isophote
sup_phi	super-coadd angle to major axis (E of N)
sup_chif_ellf	%chi-fraction for ellipse fit to super-coadd $3\sigma$ isophote
r_k20fe	20 mag/arcsec <sup>2</sup> isophotal- $K_s$ fiducial elliptical aperture semi-major axis
<band>_m_k20fe	<band> 20mag/arcsec <sup>2</sup> isophotal fiducial elliptical aperture magnitude
<band>_msig_k20fe	<band> $1\sigma$ uncertainty in <band>_m_k20fe
<band>_flg_k20fe	<band> confusion flag for <band>_m_k20fe
<band>_m_7	<band> 7'' radius circular aperture magnitude
<band>_msig_7	<band> $1\sigma$ uncertainty in 7'' circular ap. mag
<band>_flg_7	<band> confusion flag for 7'' circular ap. mag
<band>_peak	<band> peak pixel surface brightness (mag/arcs <sup>2</sup> )
<band>_5surf	<band> mean surface brightness ( $r \leq 5''$ ) (mag/arcs <sup>2</sup> )
<band>_sc_sh	<band> "shape" score; a.k.a. SH
<band>_sc_wsh	<band> "wsh" score; a.k.a. WSH
<band>_sc_r23	<band> r23 score; a.k.a. R23
<band>_bisym_rat	<band> bi-symmetric flux ratio
<band>_bisym_chi	<band> bi-symmetric cross-correlation $\chi^2$
e_score	extended source probability (1.0 = fuzziest; 2.0 = pointlike)
g_score	galaxy probability (1.0 = fuzziest; 2.0 = pointlike)

Table Notes: <band> refers to the J, H and  $K_s$  bands.

Fig 1





3

3

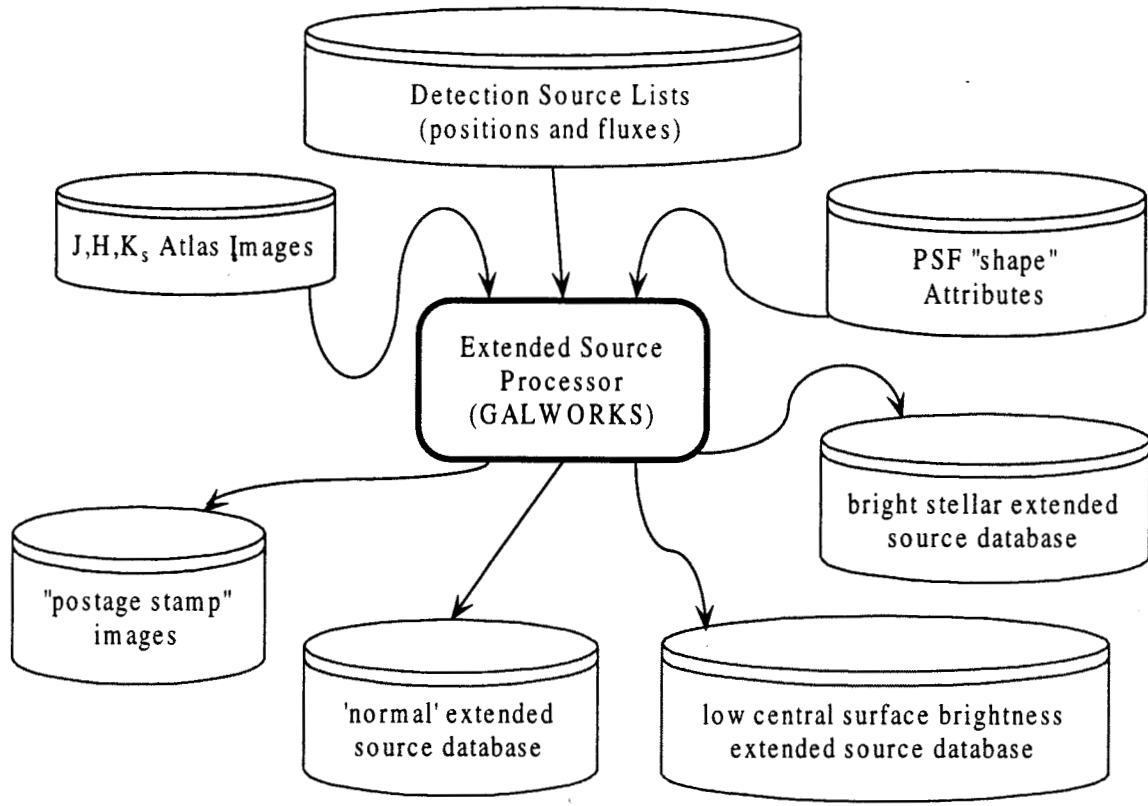


Fig 4

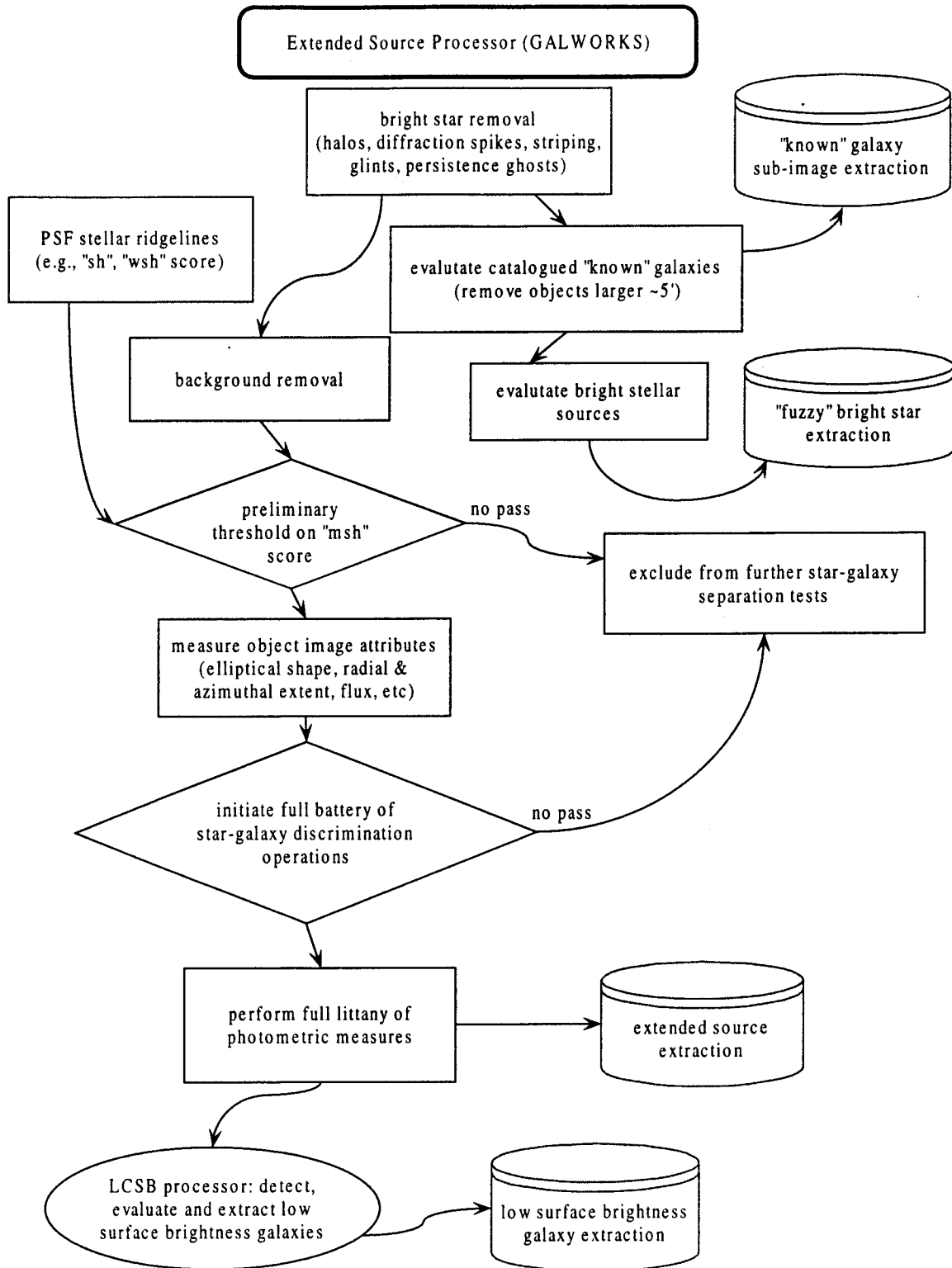
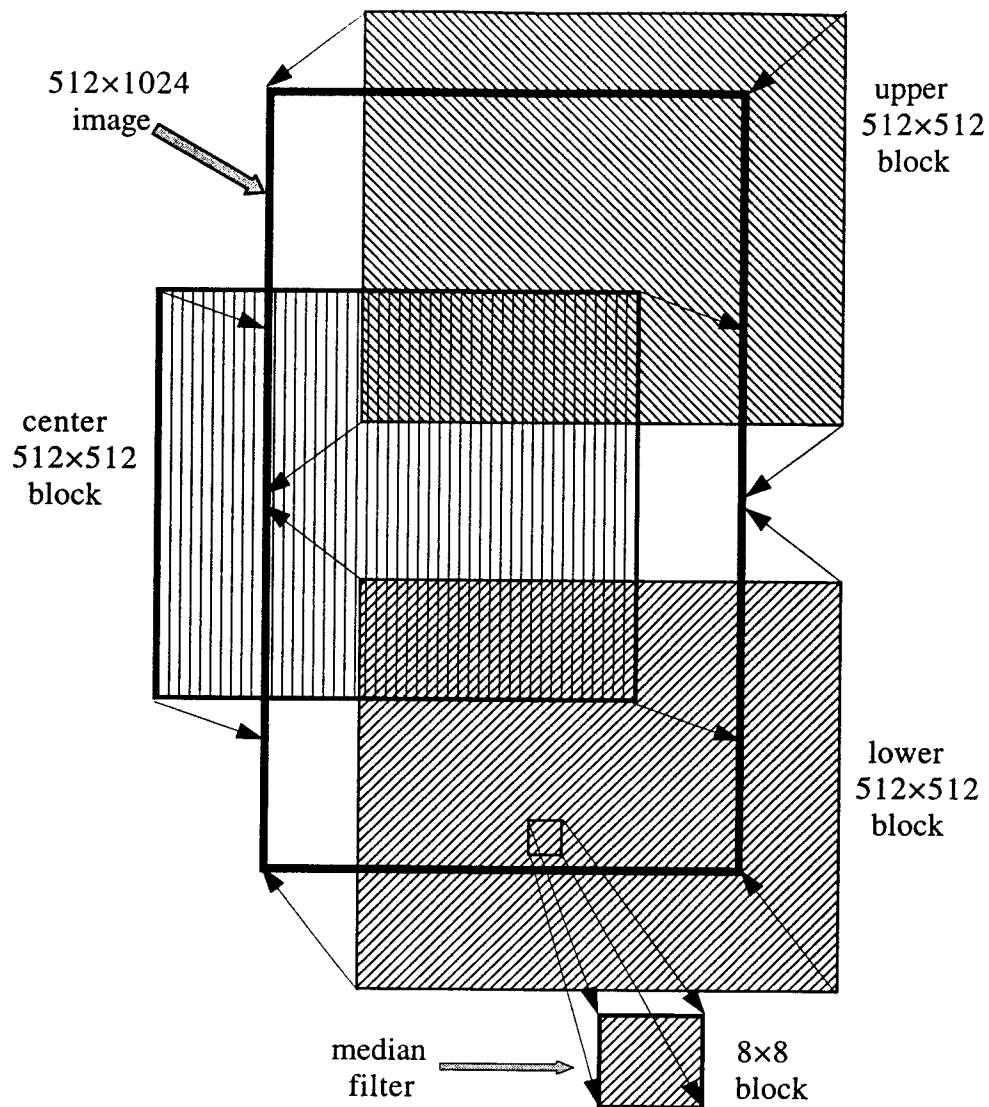
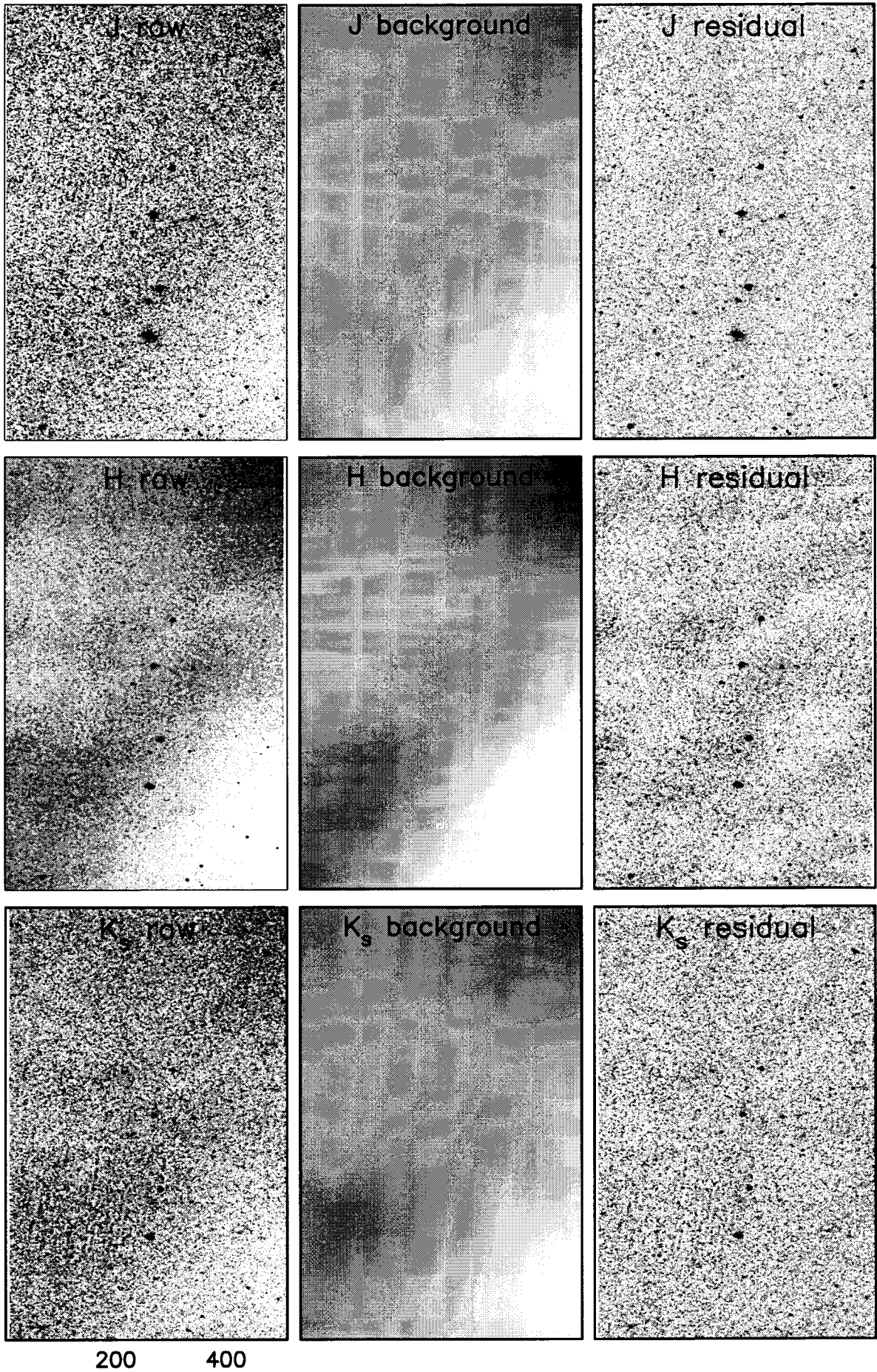


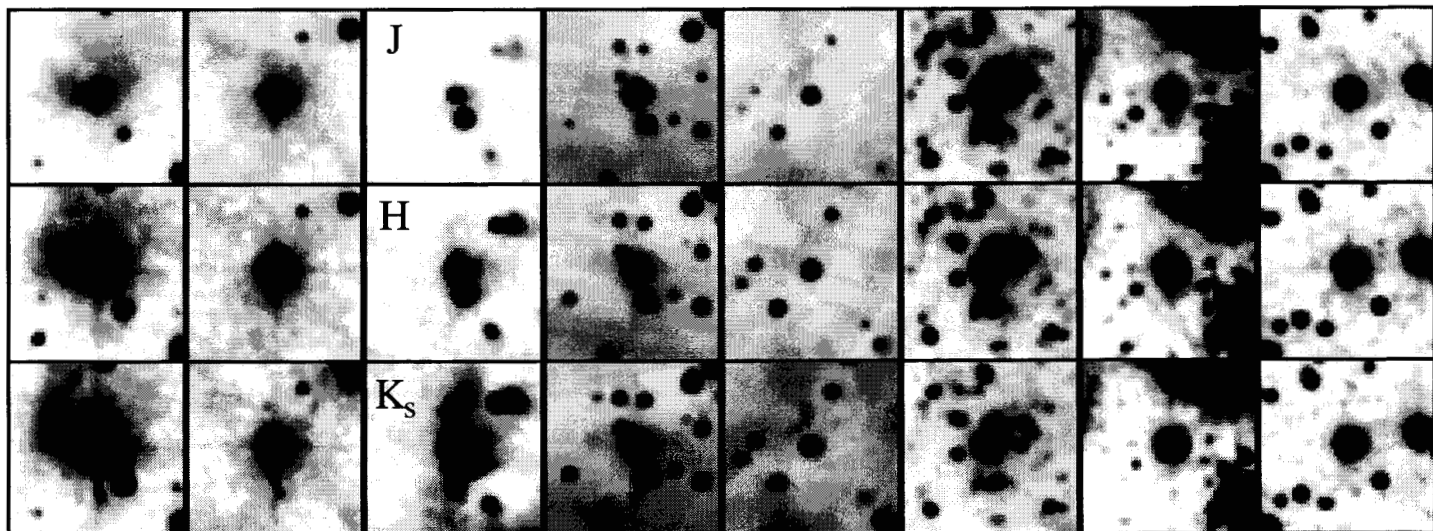
Fig 5.

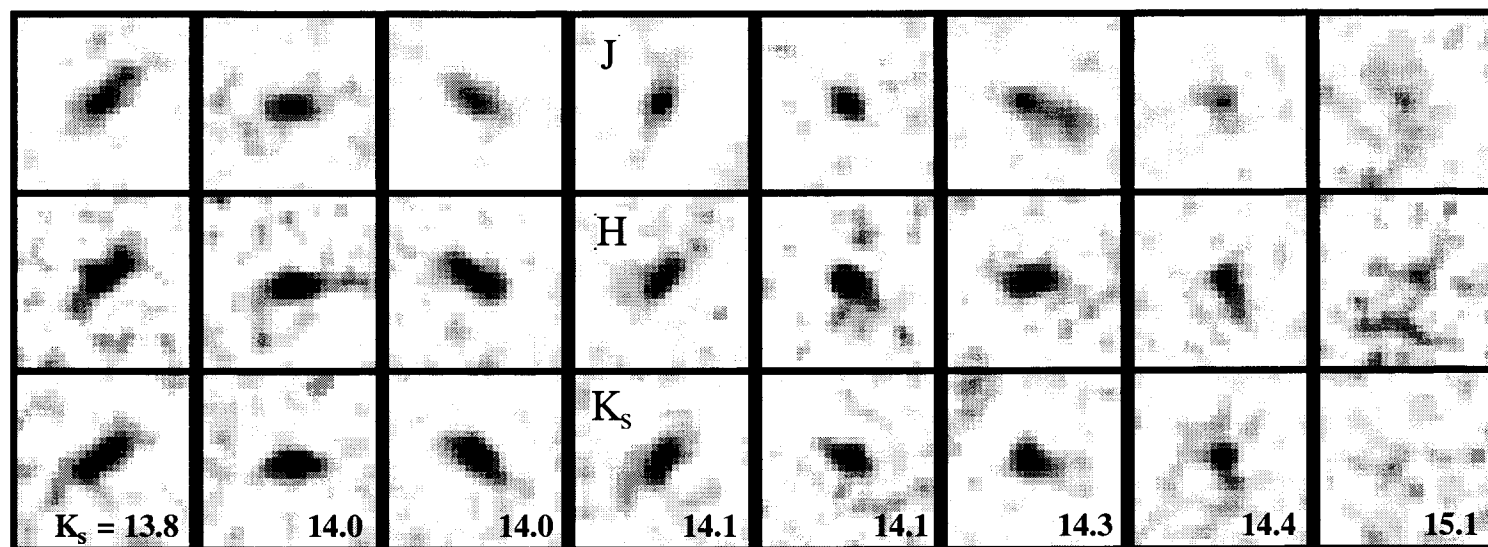






← Galactic Sources from Orion → ← Galactic Sources from LMC





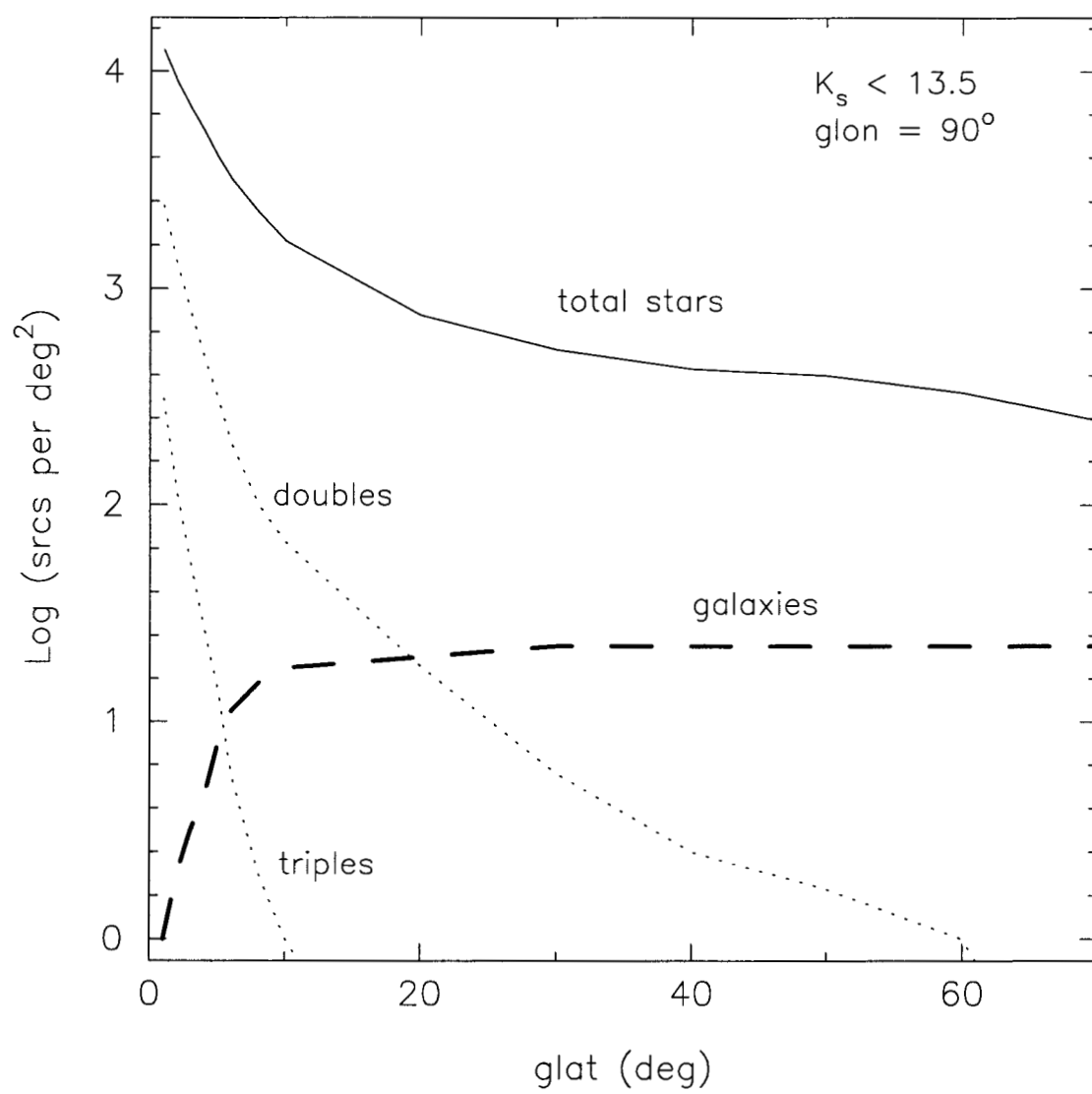


Fig 10

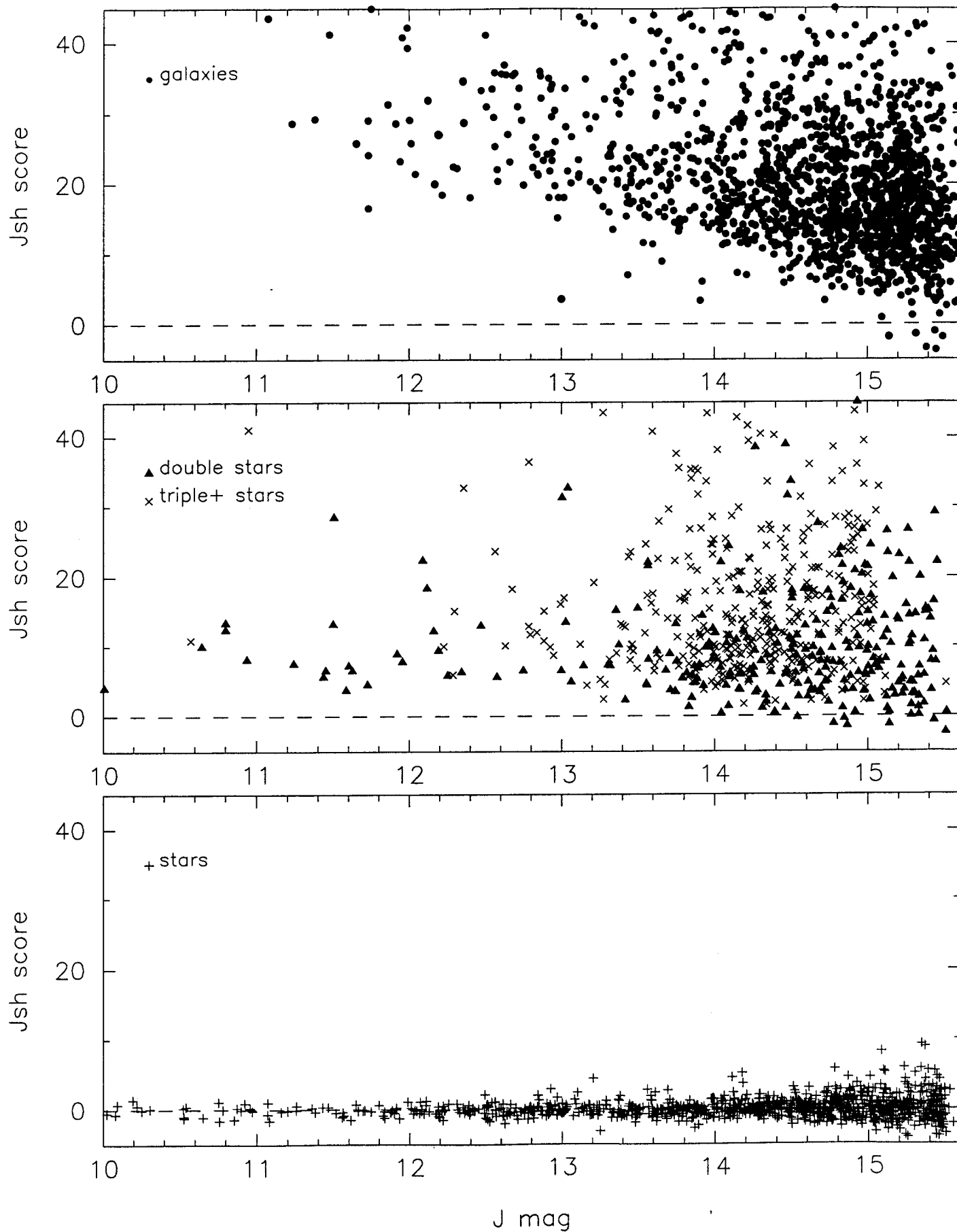


FIG 11

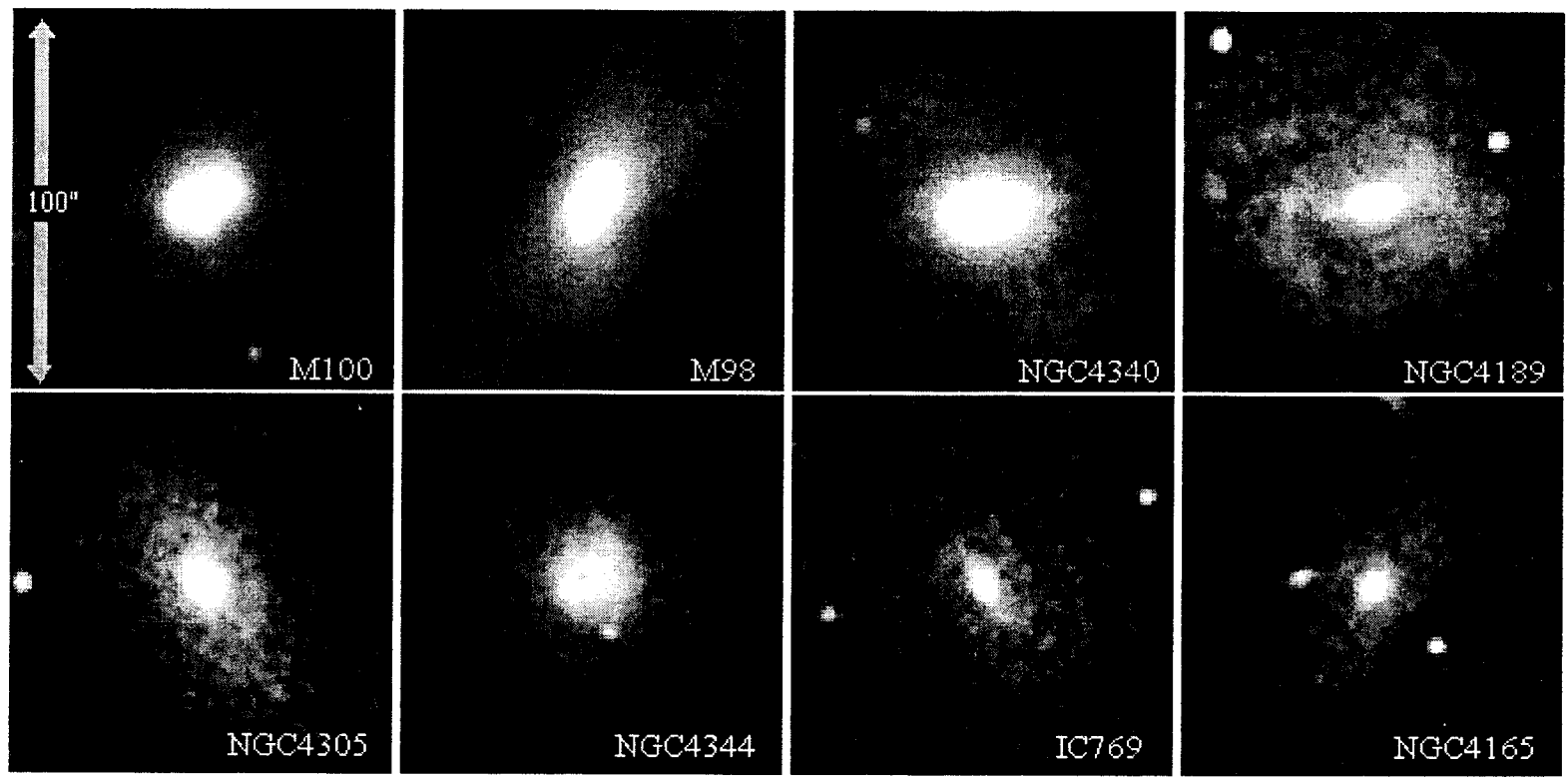
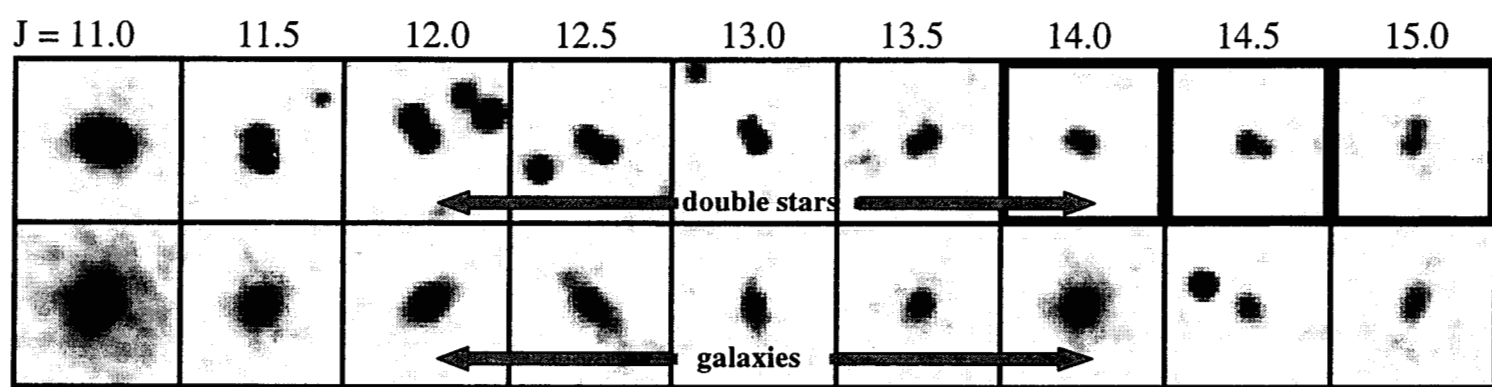
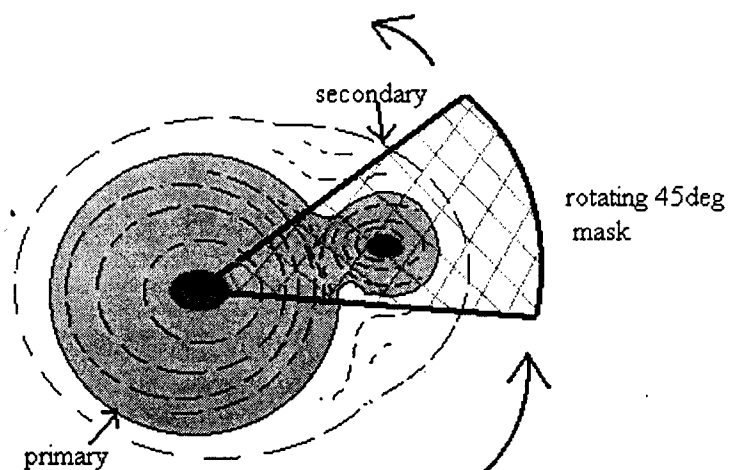
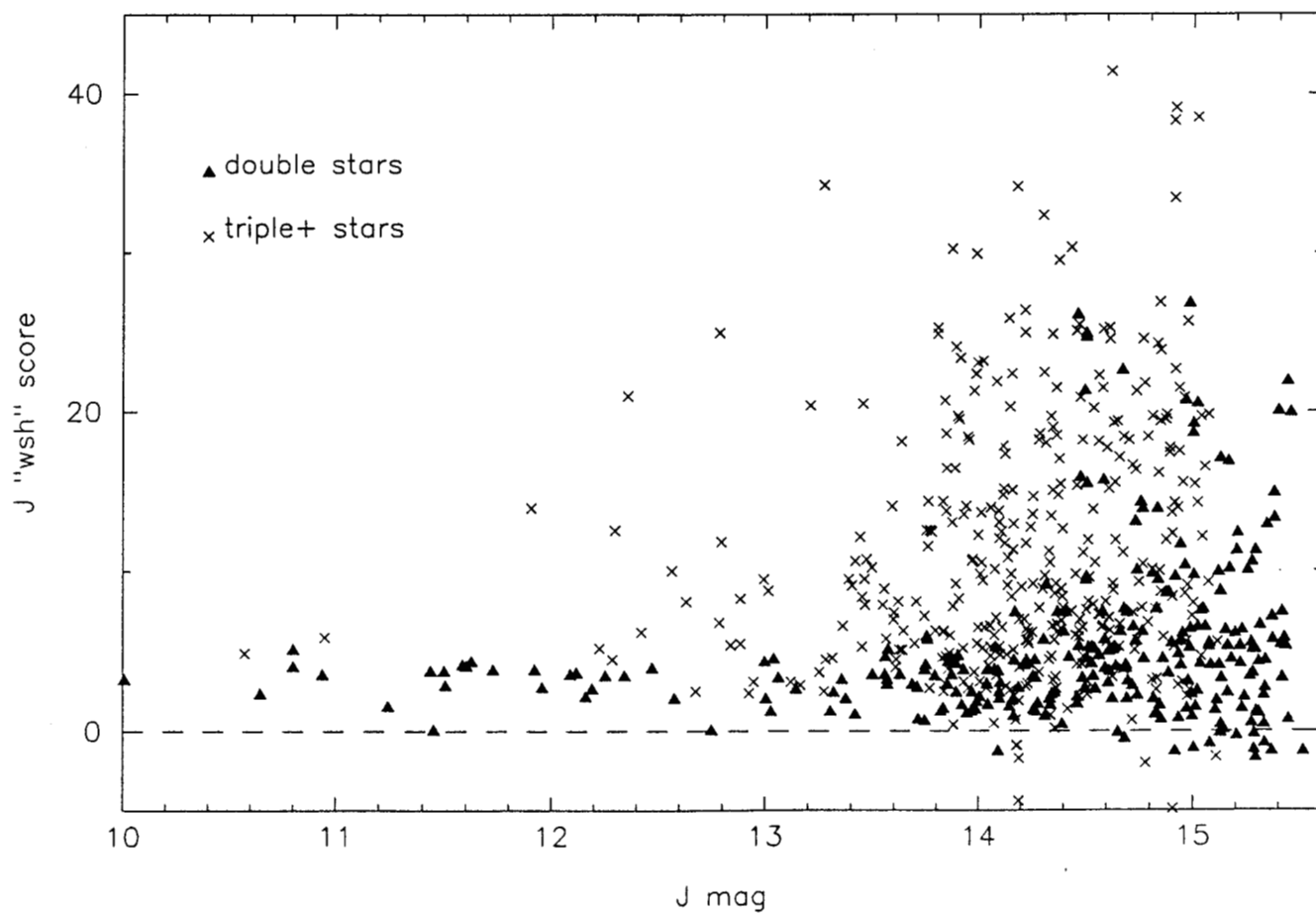
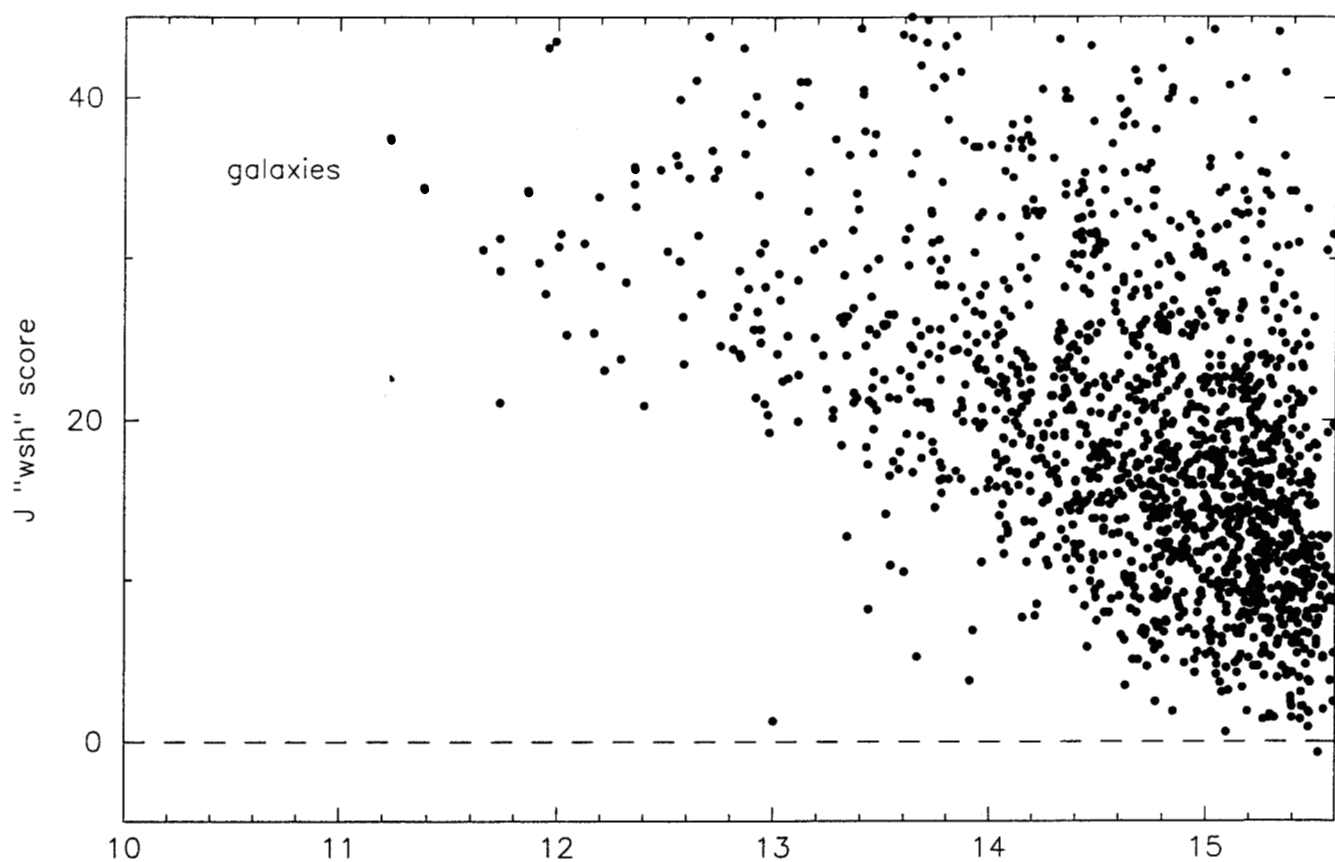


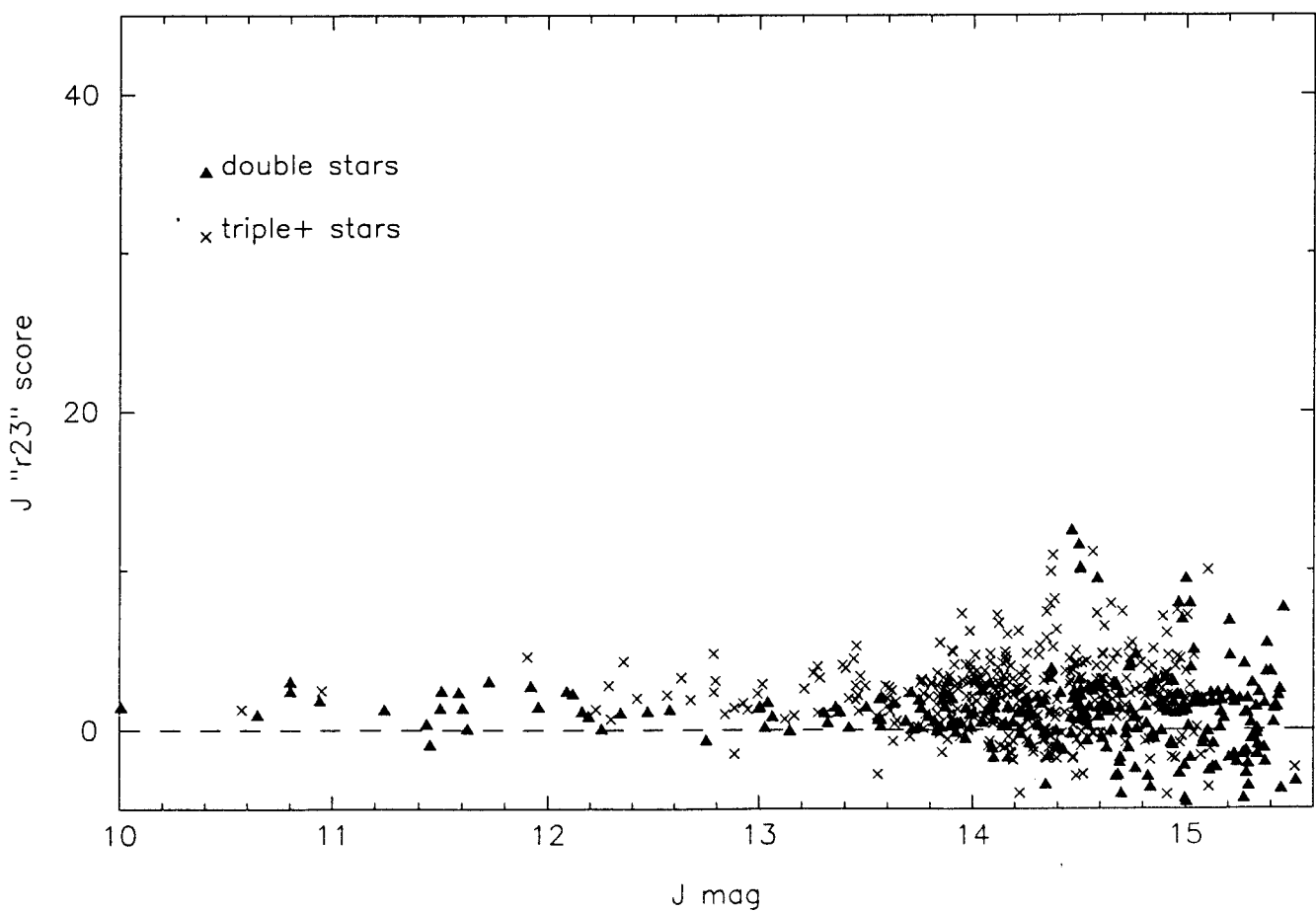
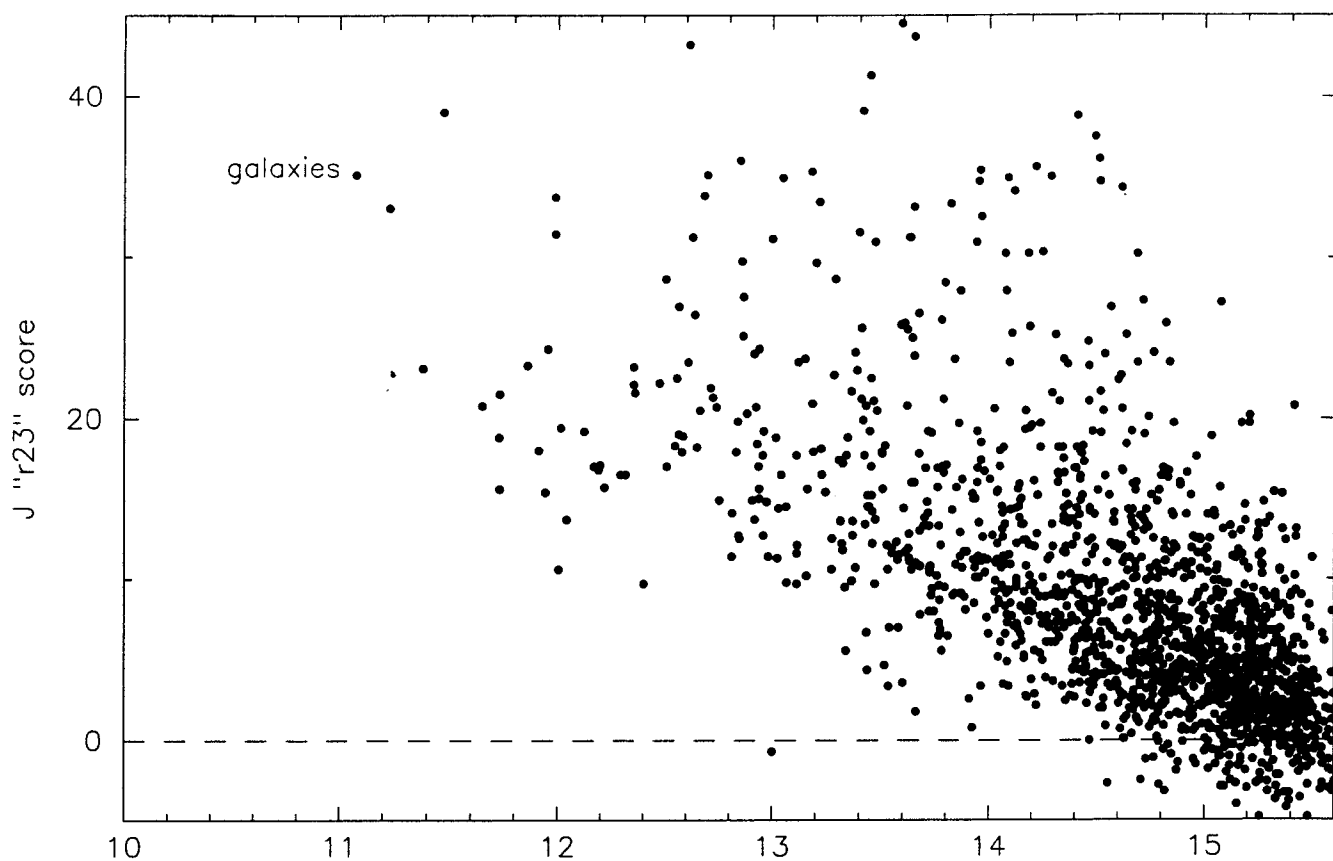
Fig 12

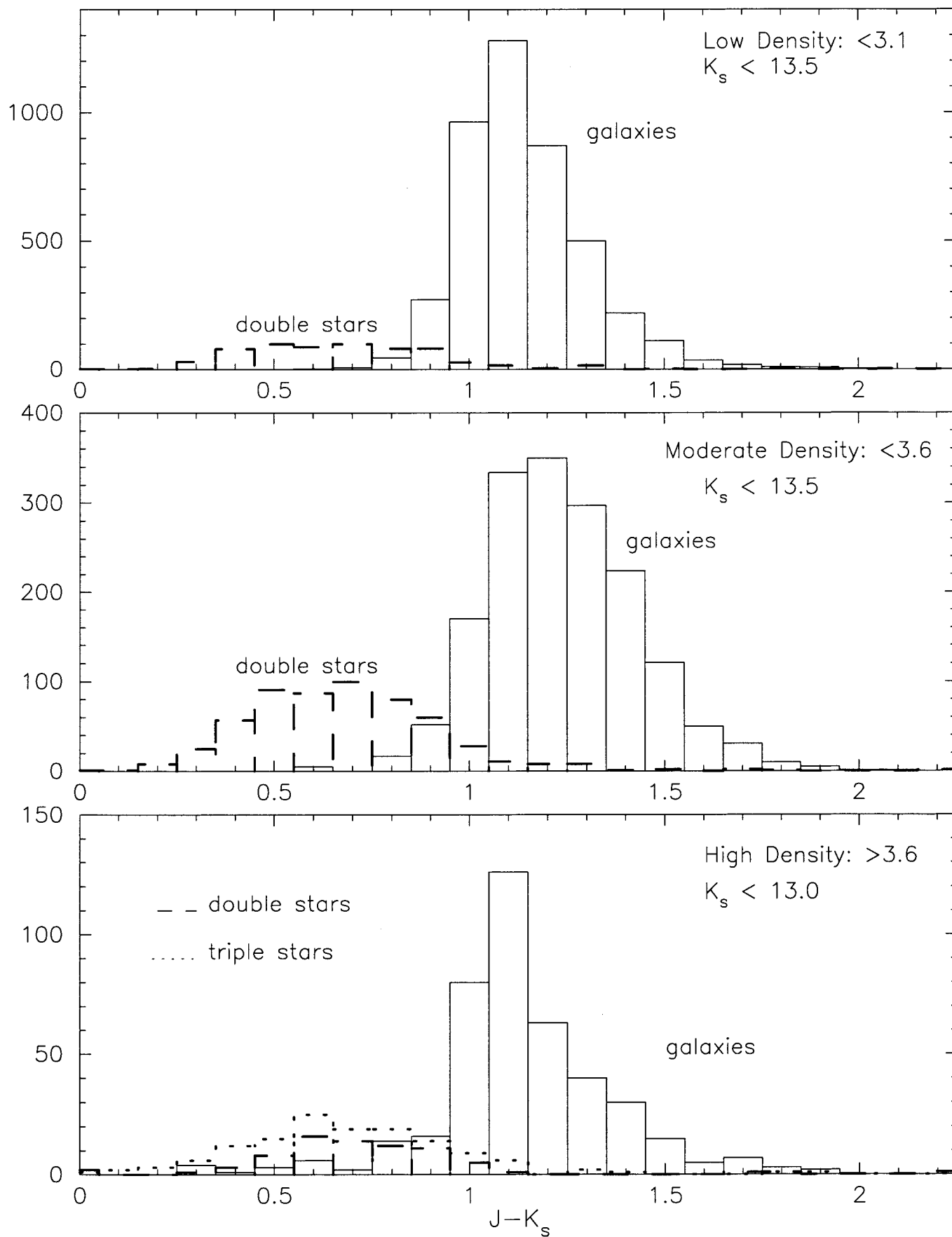


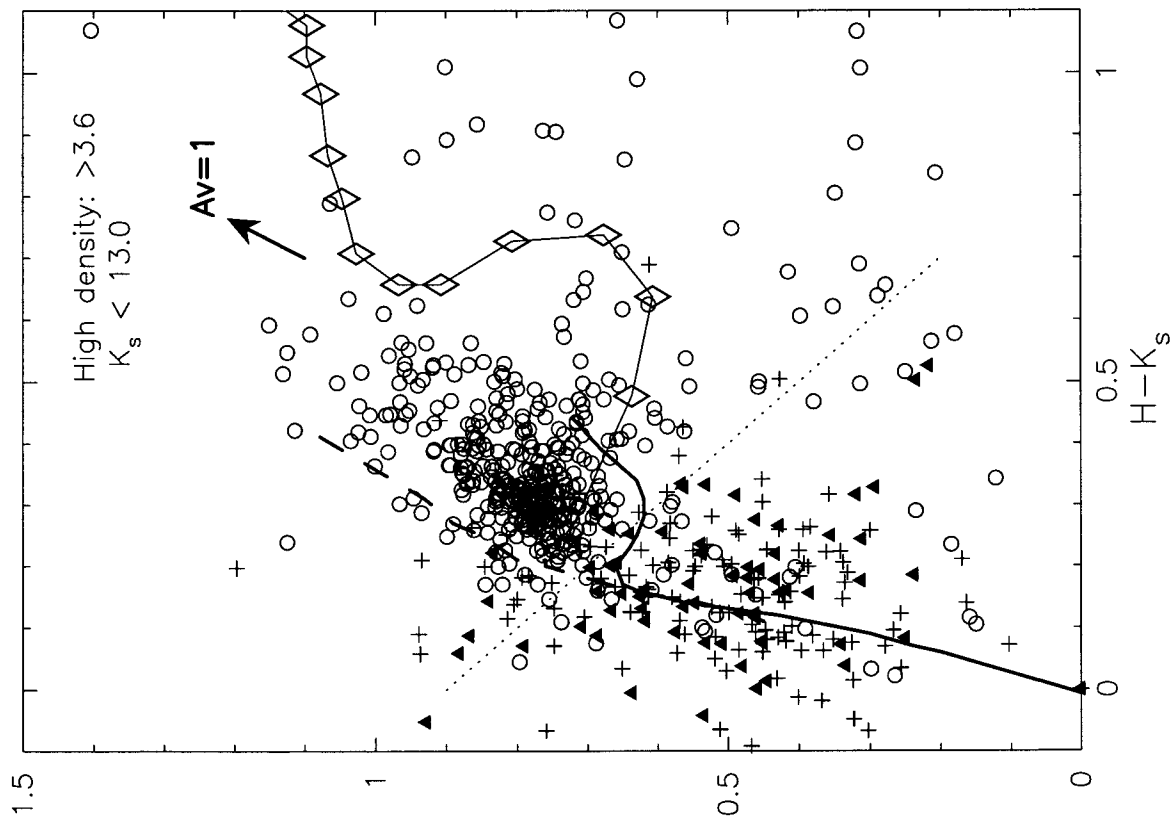
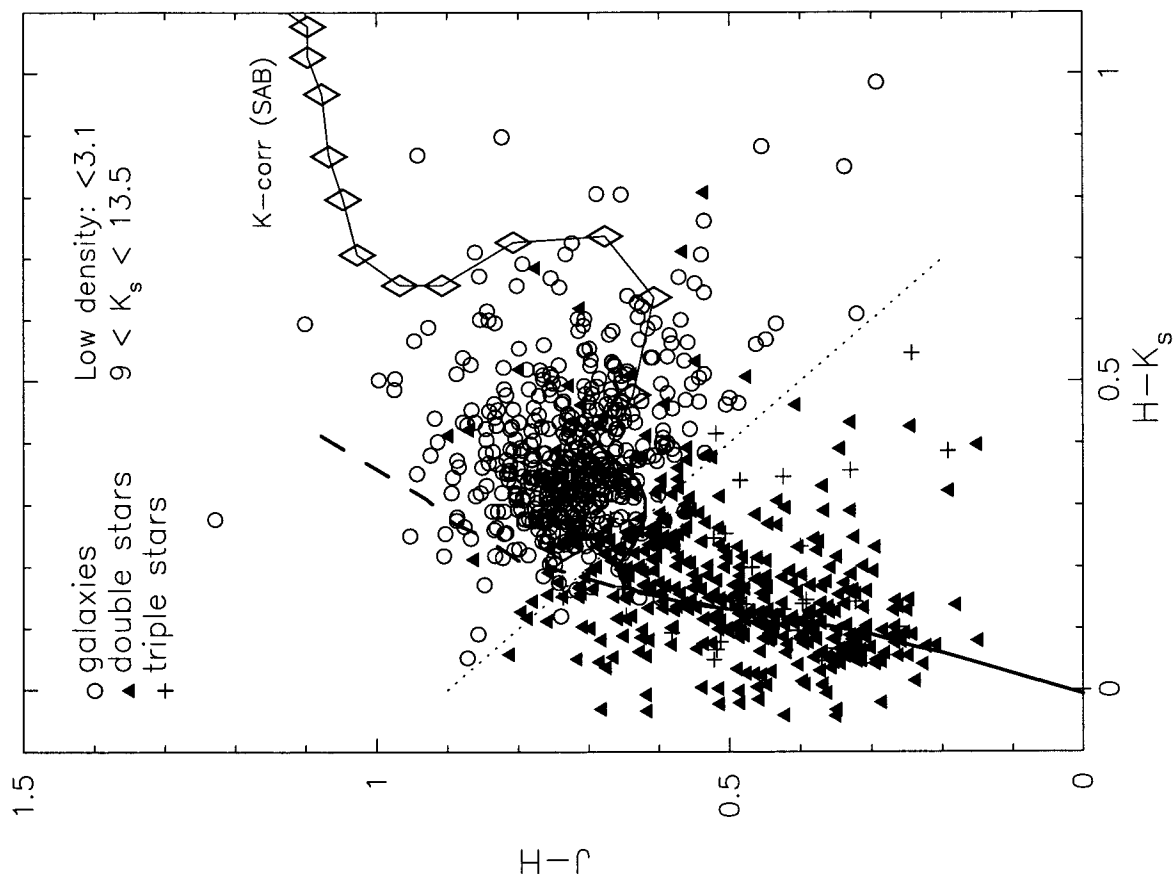


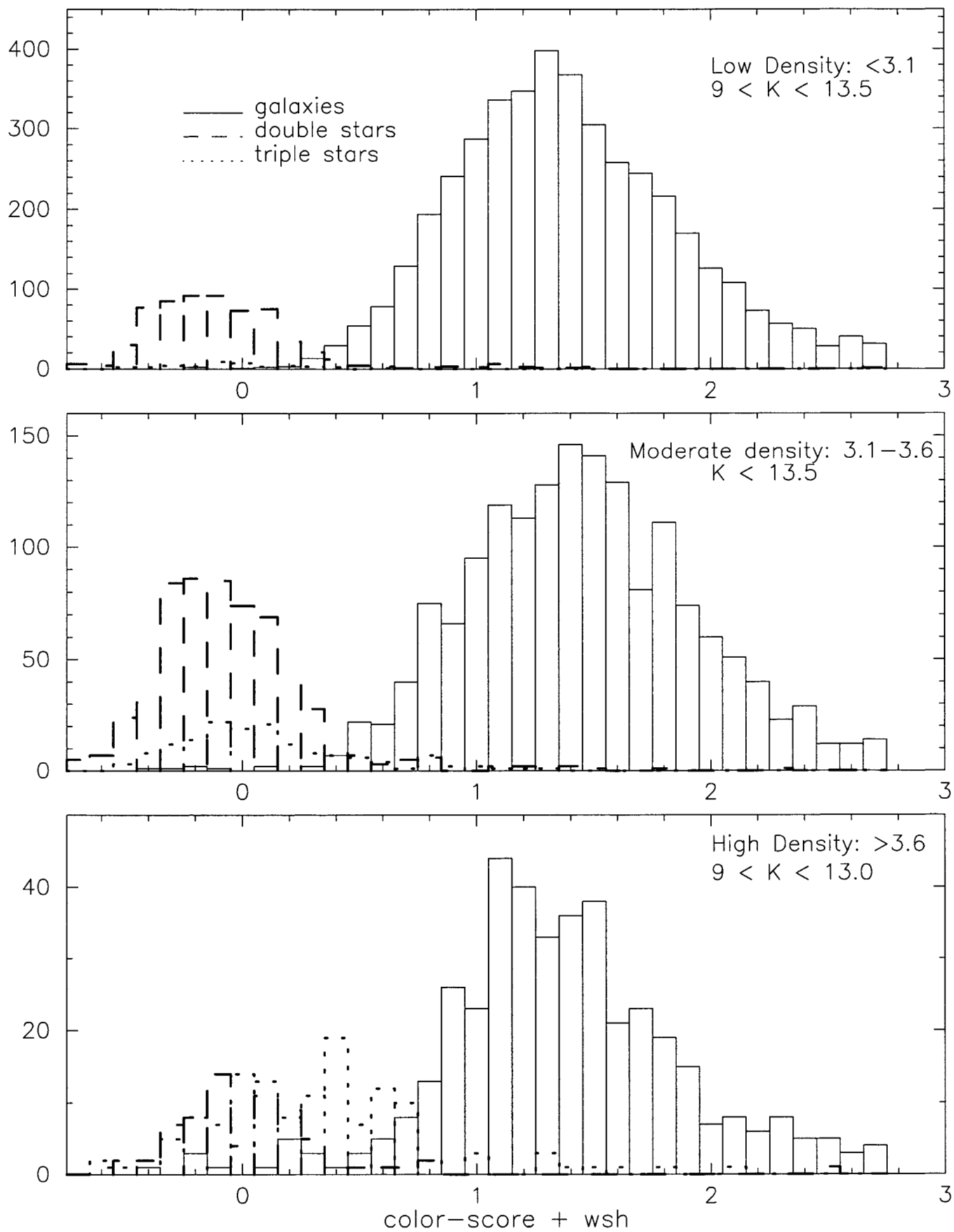


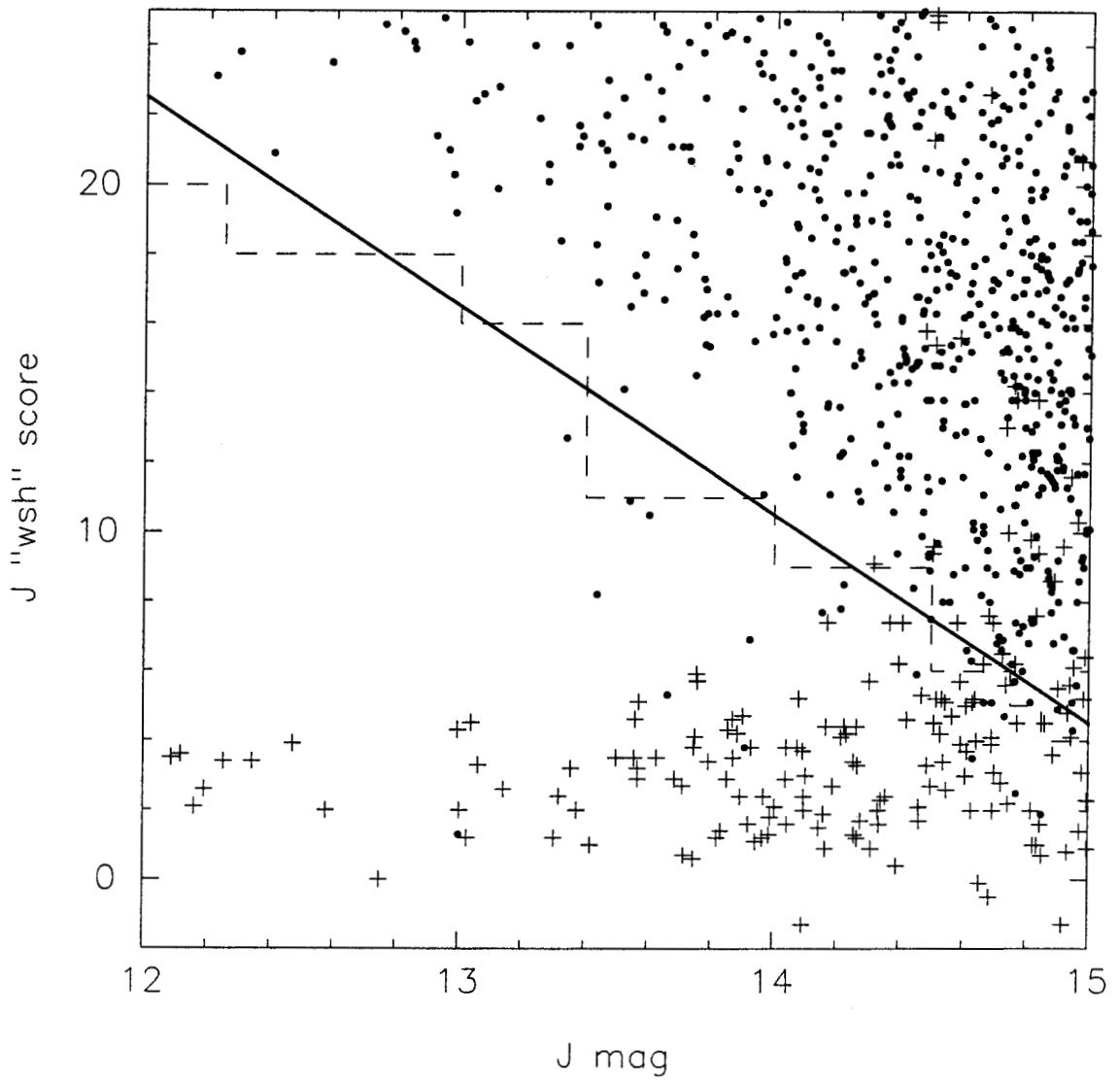


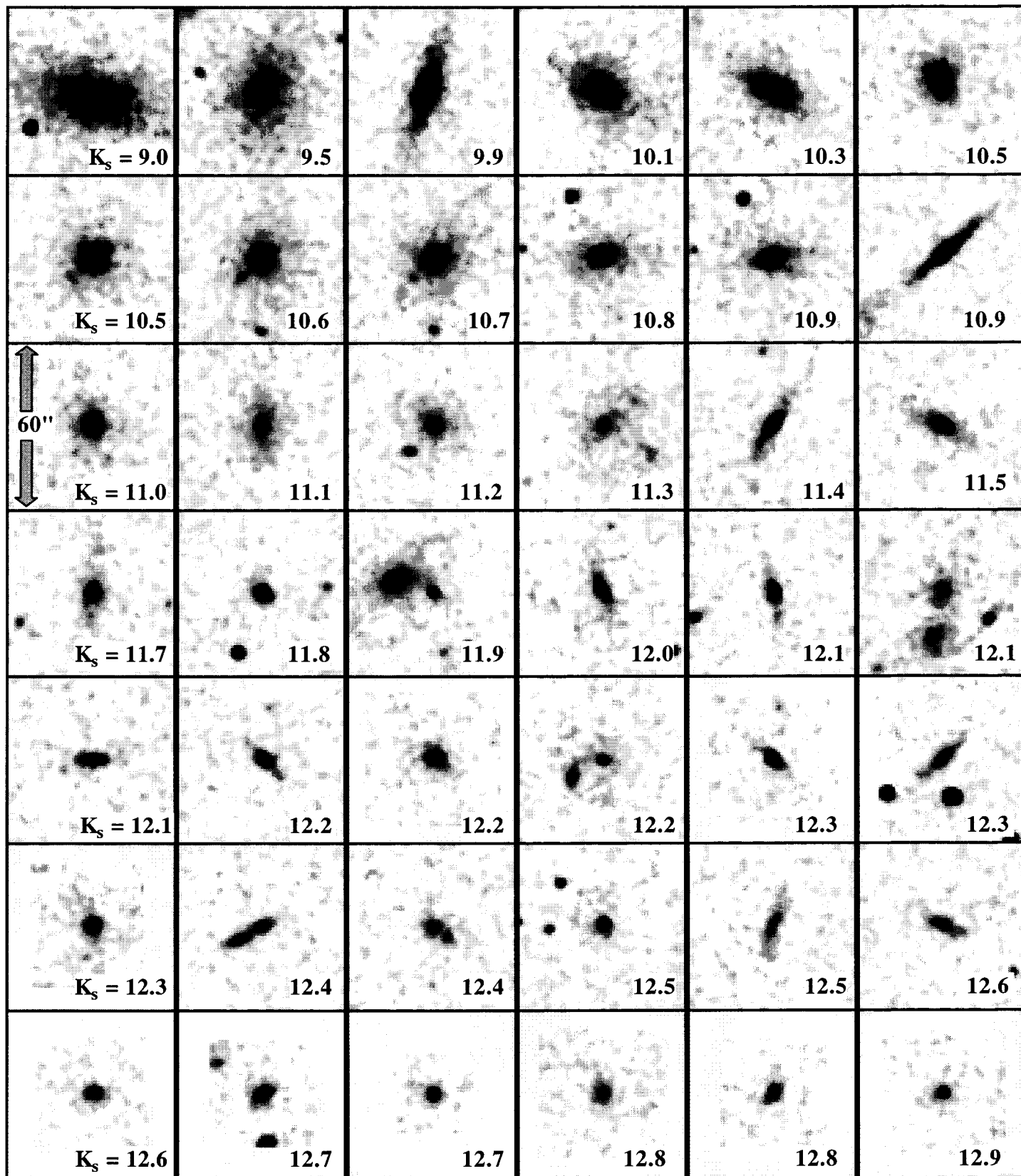


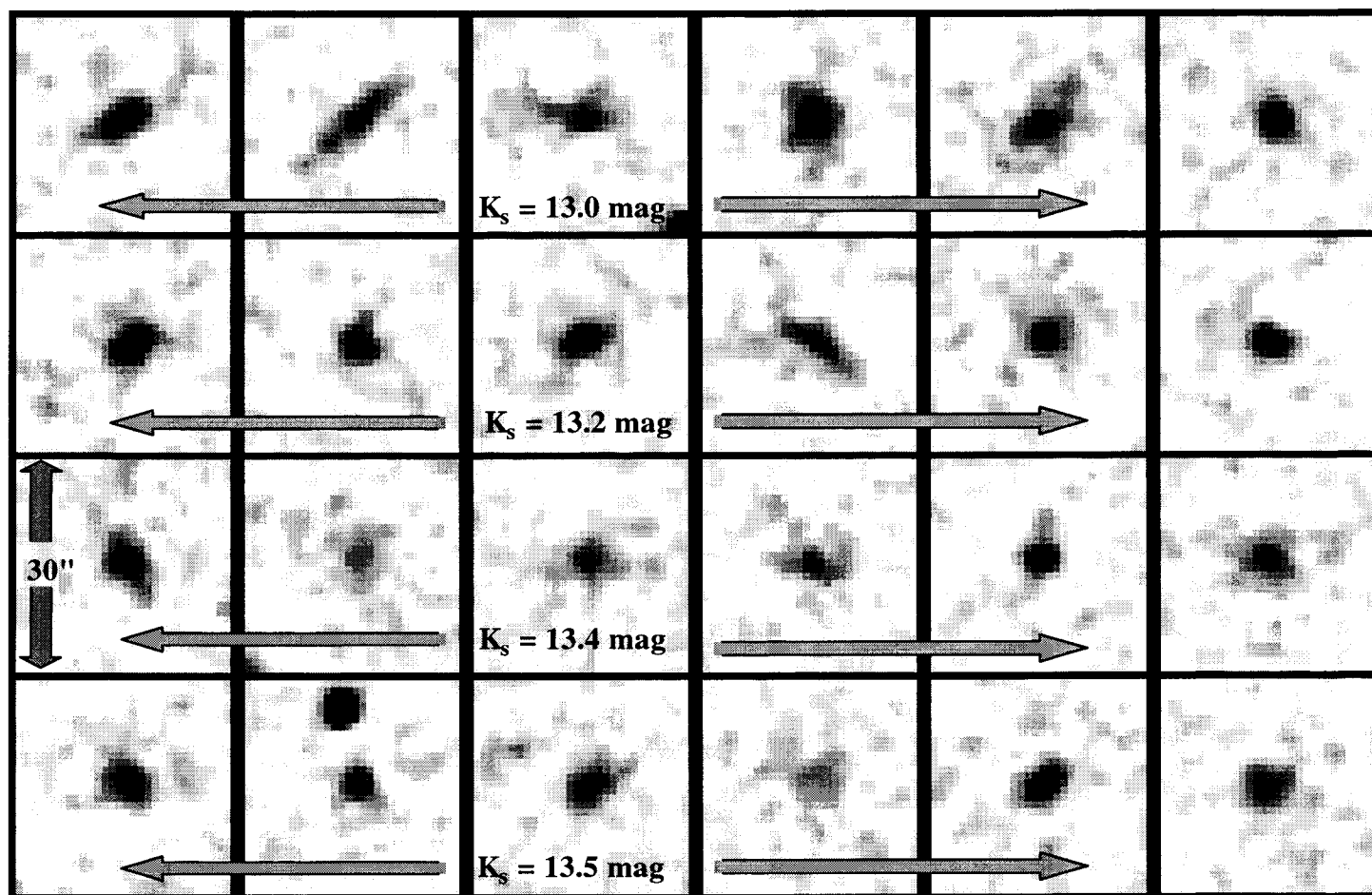


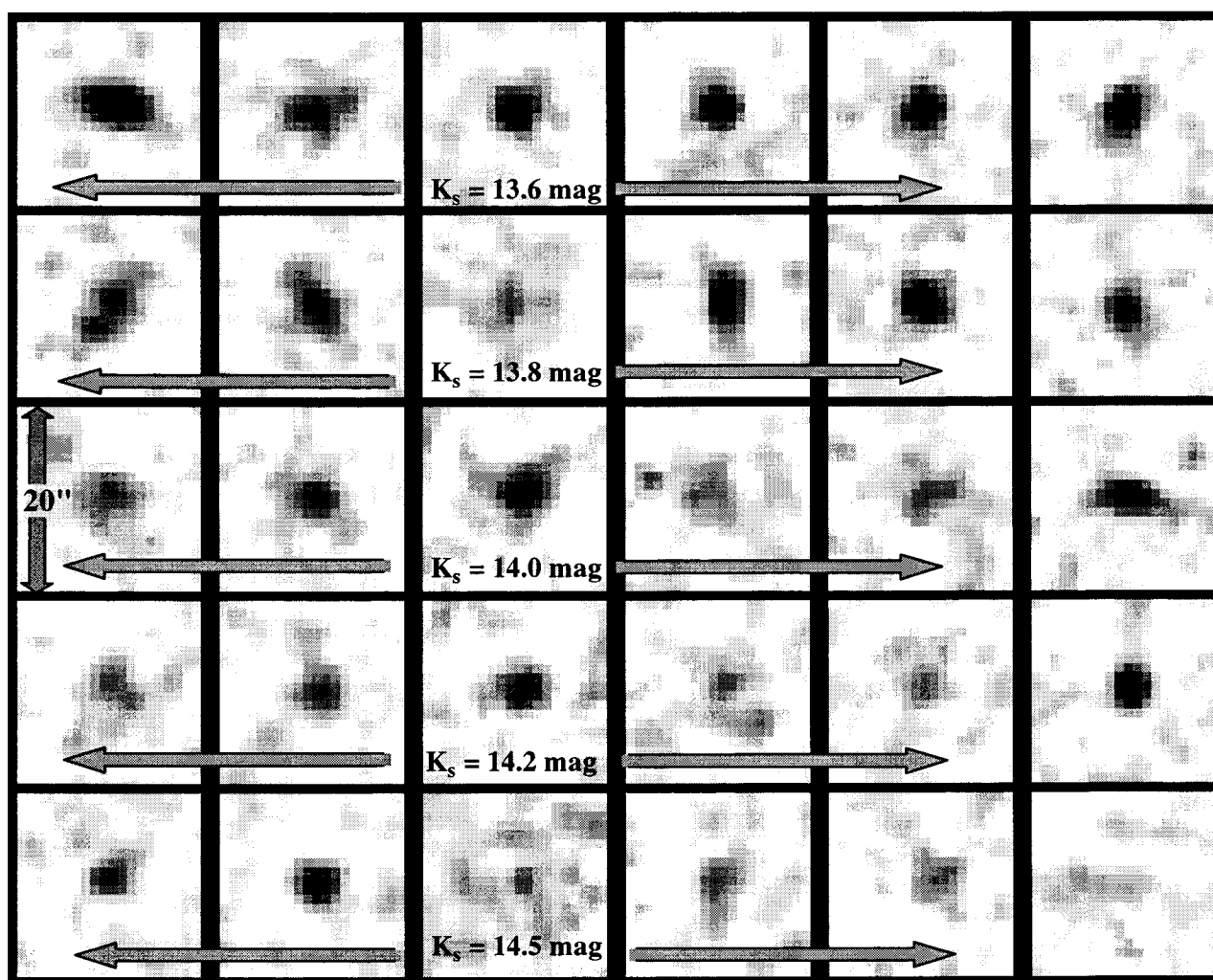






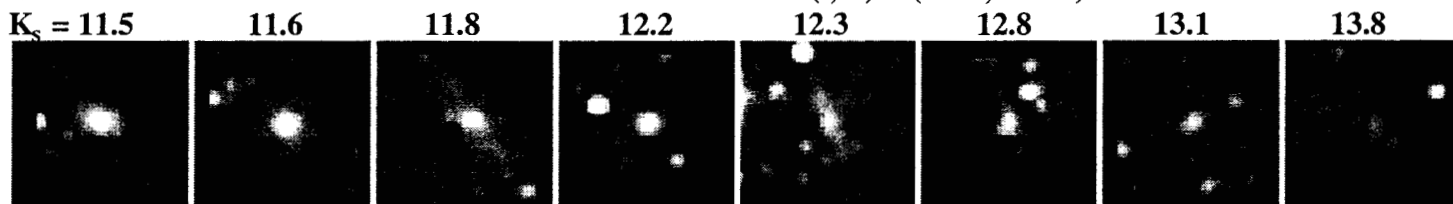








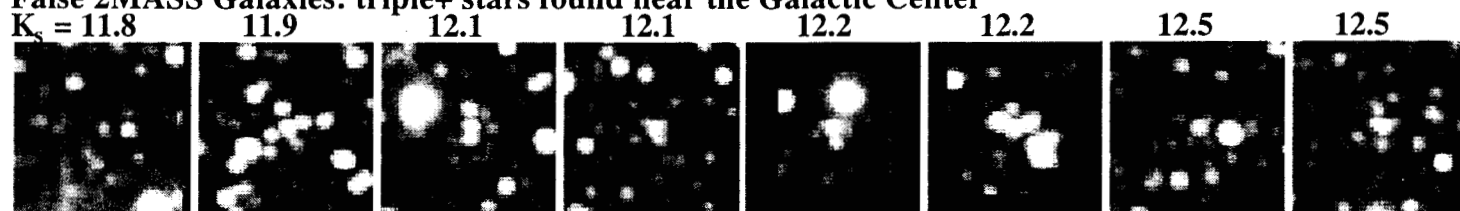
**Galaxies & Extended Sources in the Zone of Avoidance:  $(l, b) = (240^\circ, +4.5^\circ)$**



**Galaxies & Extended Sources in the Galactic Bulge:  $(l, b) = (12^\circ, +5^\circ)$**



**False 2MASS Galaxies: triple+ stars found near the Galactic Center**



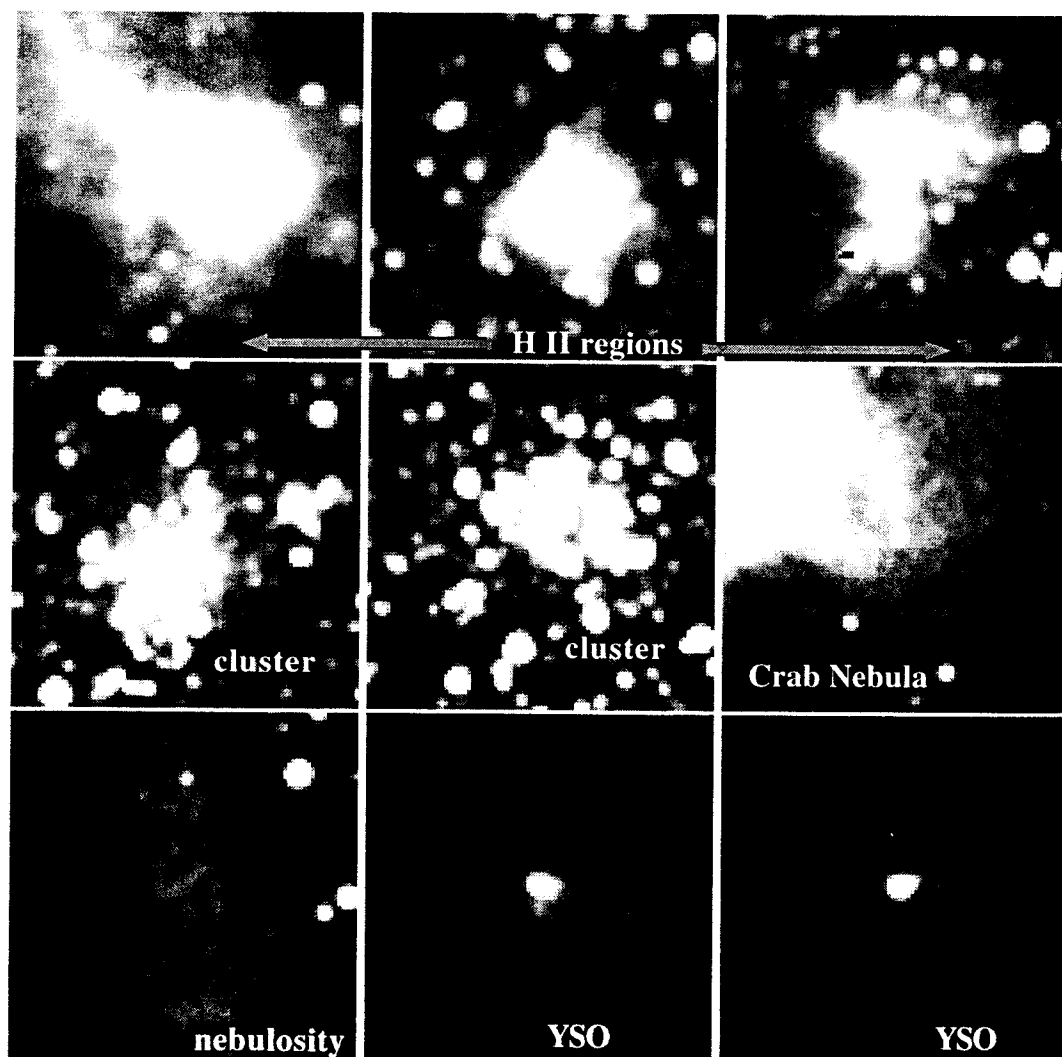


Fig 25

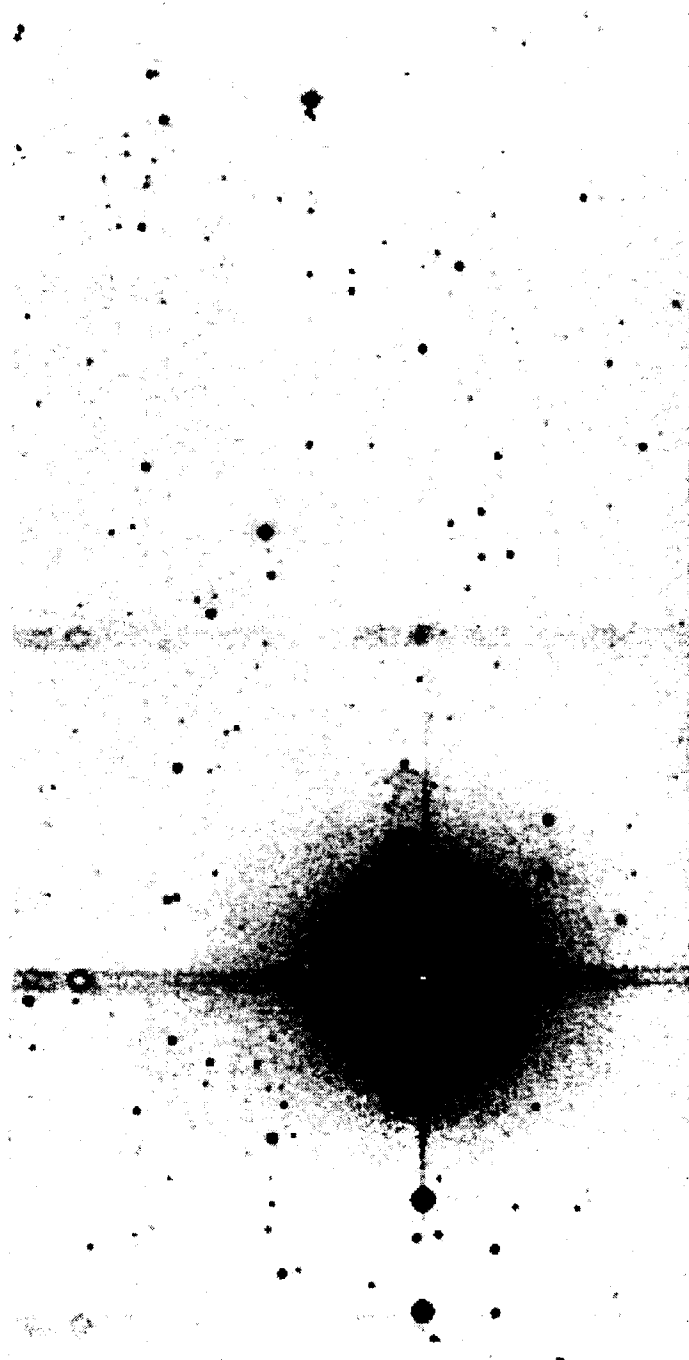
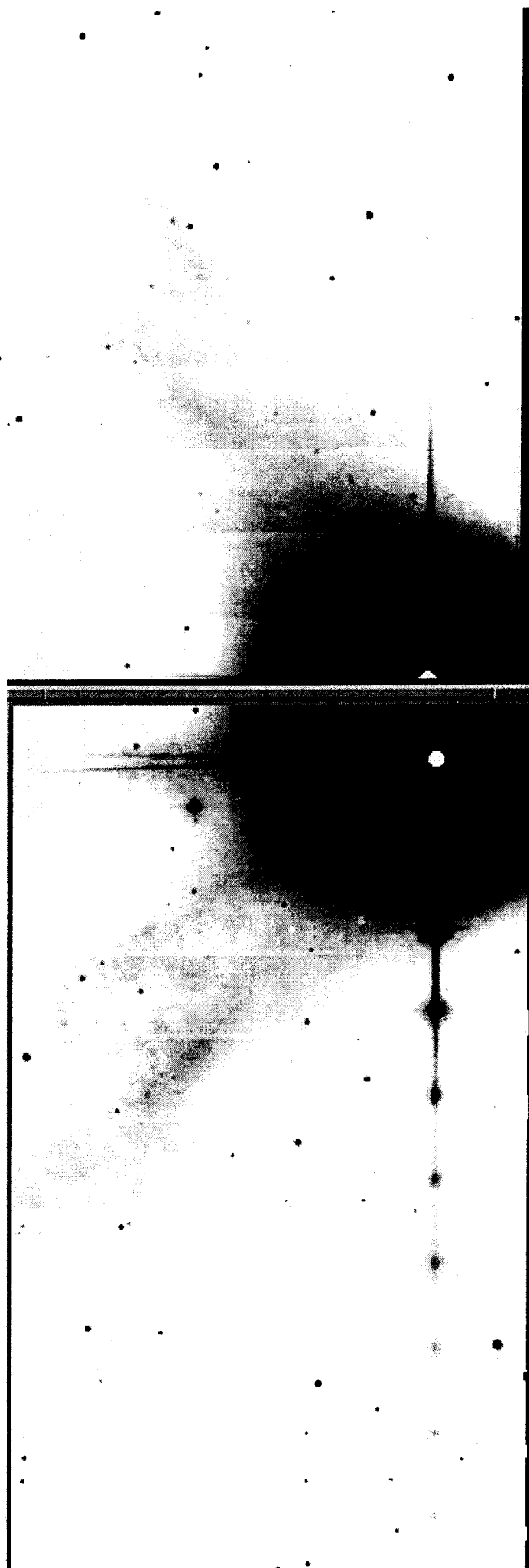
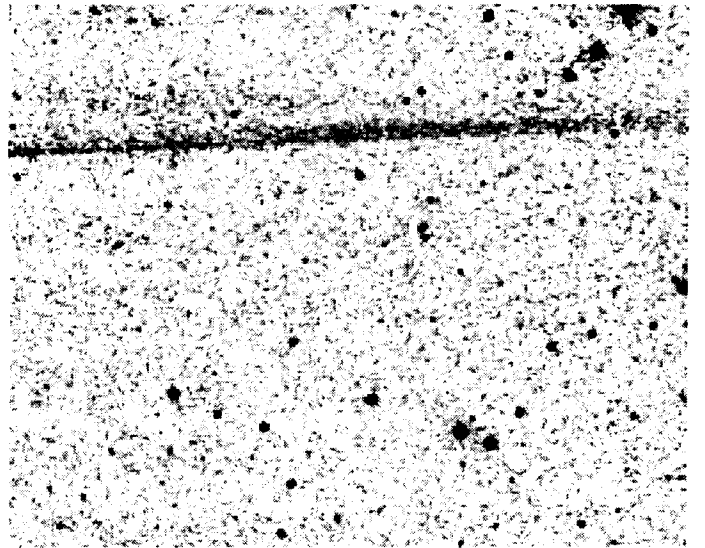


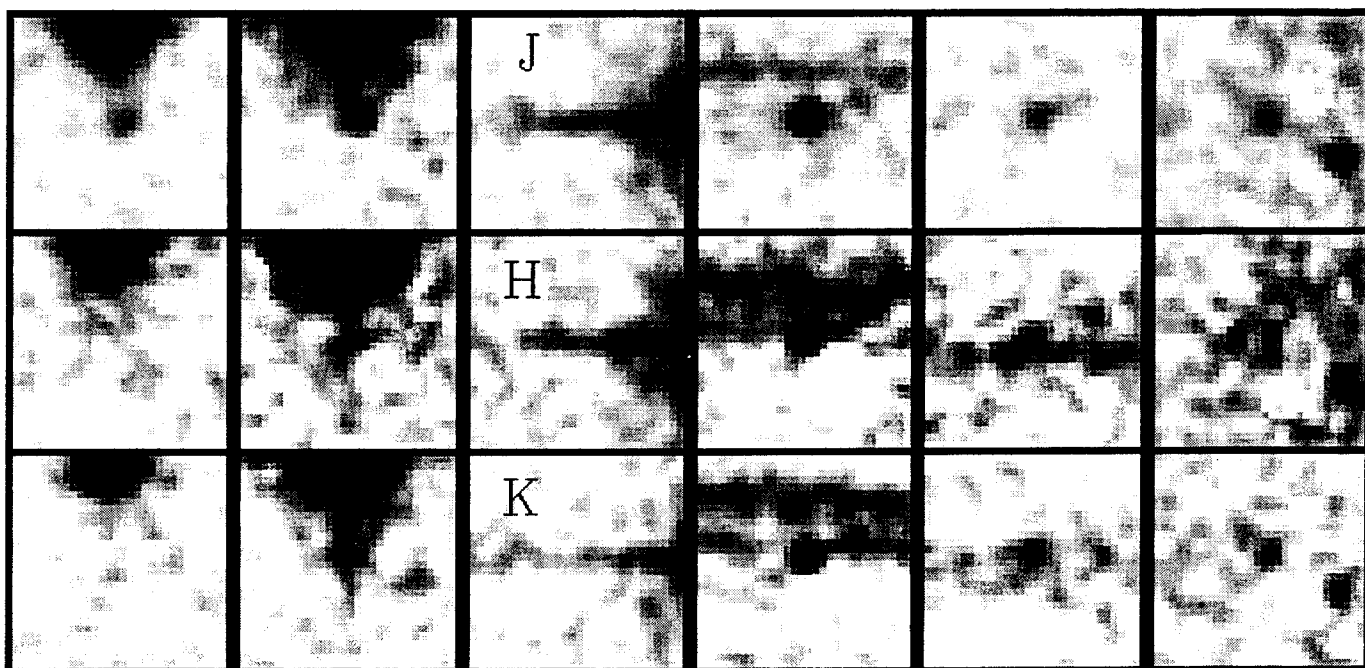
Fig 26

Fig 26



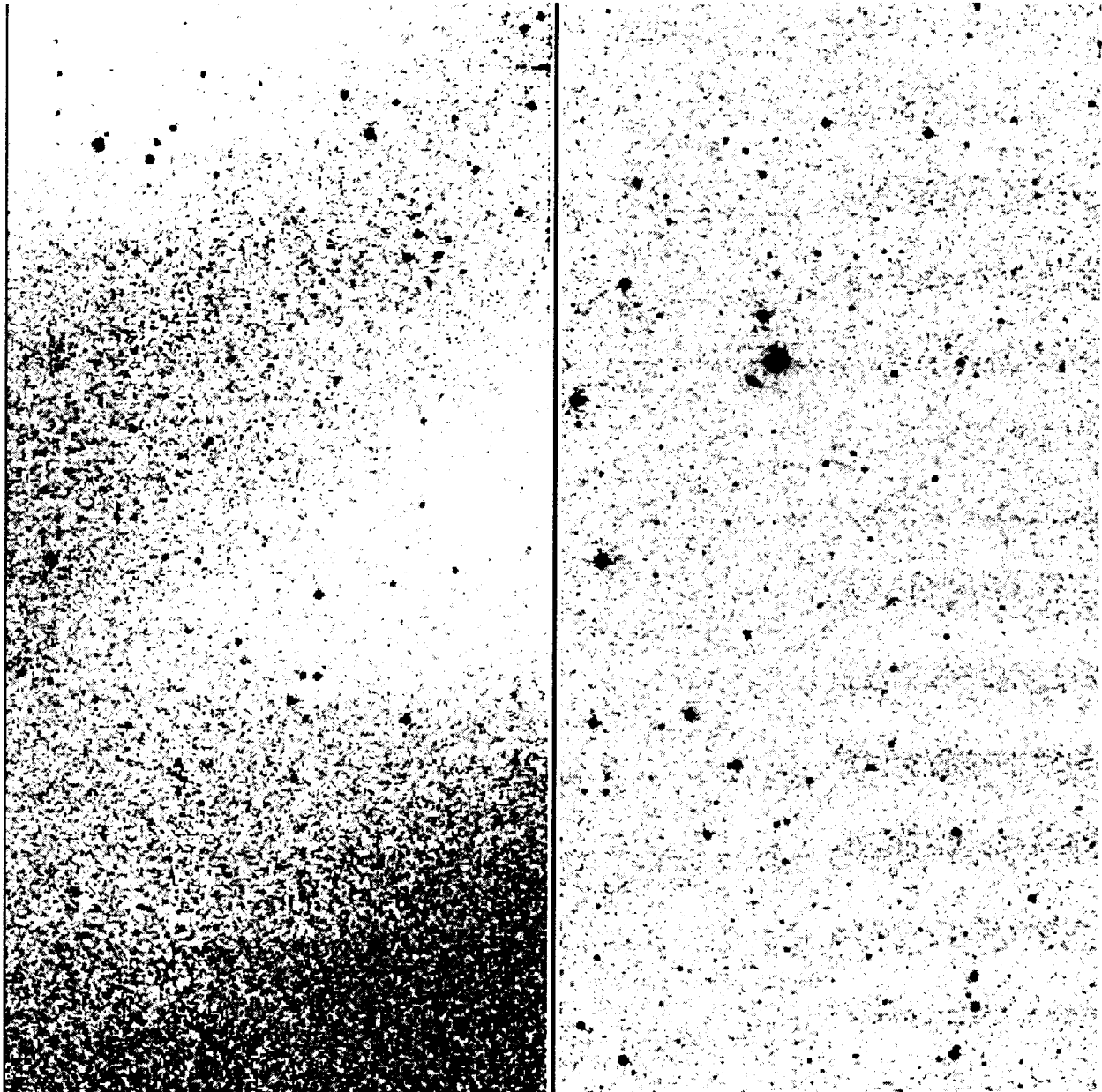
27





804

Fig A.1



A1

Fig C.1

